

# SEARCH REQUEST FORM

Requestor's  
Name: \_\_\_\_\_

Serial  
Number: \_\_\_\_\_

Date: \_\_\_\_\_

Phone: \_\_\_\_\_

Art Unit: \_\_\_\_\_

## Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

## STAFF USE ONLY

Date completed: 04-12-04  
Searcher: Belding C 2528  
Terminal time: 25  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 25  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

### Search Site

\_\_\_\_\_ STIC  
\_\_\_\_\_ CM-1  
\_\_\_\_\_ Pre-S

### Type of Search

\_\_\_\_\_ N.A. Sequence  
\_\_\_\_\_ A.A. Sequence  
\_\_\_\_\_ Structure  
\_\_\_\_\_ Bibliographic

### Vendors

\_\_\_\_\_ IG  
\_\_\_\_\_ STN  
\_\_\_\_\_ Dialog  
\_\_\_\_\_ APS  
\_\_\_\_\_ Geninfo  
\_\_\_\_\_ SDC  
\_\_\_\_\_ DARC/Questel  
\_\_\_\_\_ Other Geninfo

This Page Blank (uspto)

STIC-Biotech/ChemLib

118960

From: Leffers, Gerald  
Sent: Wednesday, April 07, 2004 1:37 PM  
To: STIC-Biotech/ChemLib  
Subject: 09/430,590

Please interference search/search SEQ ID NO: 3 (~ 6.5 kb). Thank you. Gerry Leffers

*Gerald G. Leffers Jr., PhD*  
Primary Examiner, Art Unit 1636  
Remsen Building, Room 02A69  
(571) 272-0772

RECEIVED  
APR - 7 2004  
STIC

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:  
NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 08:06:01 ; Search time 15972 Seconds

(without alignments)  
17438.163 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426

Sequence: 1 tcttggtctgtgcactattc.....agaagttatattccatca 6426

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

GenBank1:  
1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_pl:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_hcg\_hum:\*  
31: em\_hcg\_in:\*  
32: em\_hcg\_om:\*  
33: em\_hcg\_mus:\*  
34: em\_hcg\_pln:\*  
35: em\_hcg\_rtd:\*  
36: em\_hcg\_mam:\*  
37: em\_hcg\_yrt:\*  
38: em\_hcg\_hum:\*  
39: em\_hcg\_mus:\*  
40: em\_hcg\_om:\*  
41: em\_hcg\_other:\*

Score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6426	100.0	6426	8 AF007776	AF007776 Candida a
2	6421.2	99.9	6980	8 AF050215	AF050215 Candida a
3	432	6.7	4872	8 CAL251464	AJ251464 Candida a
4	400	6.2	408	8 AF030556	AF030556 Candida a
5	363.8	5.7	5928	8 DHA439551	AJ439551 Debaryomyces
6	339.2	5.3	1470	8 AF078809	AF078809 Candida a
7	166	2.6	1888	6 AX073190	AX073190 Sequence
8	107	1.7	110000	2 PFMA113_07	Continuation (8 of
9	105.6	1.6	128101	2 AC140550	AC140550 Medicago
10	105.6	1.6	146275	2 AL935272	AL935272 Danio rer
11	105.4	1.6	72880	8 AP006375	AP006375 Lotus cor
12	105.4	1.6	127731	8 AP006094	AP006094 Lotus cor
13	103.8	1.6	105887	8 AP006381	AP006381 Lotus cor
14	103.4	1.6	129571	2 AC137081	AC137081 Medicago
15	102.2	1.6	134844	2 AC135386	AC135386 Medicago
16	101	1.6	82726	2 AC137509	AC137509 Medicago
17	101	1.6	91075	2 AC140772	AC140772 Medicago
18	101	1.6	106730	8 AC144724	AC144724 Medicago
19	101	1.6	127857	2 AC133780	AC133780 Medicago
20	100.4	1.6	116807	2 BX890614	BX890614 Danio rer
21	99.8	1.6	105246	8 AC139747	AC139747 Medicago
22	99.8	1.6	131746	2 AC137553	AC137553 Medicago
23	99.4	1.5	80561	8 AC073555	AC073555 Arabidops
24	99.4	1.5	115995	8 AC146566	AC146566 Medicago
25	98.8	1.5	109155	8 ATT20K12	AL177898 Arabidops
26	98.8	1.5	133299	8 OSUN00098	AL60658 Oryza sat
27	98	1.5	85661	8 AC079131	AC079131 Arabidops
28	98	1.5	103637	8 AC079604	AC079604 Arabidops
29	97.4	1.5	133397	2 AC147008	AC147008 Medicago
30	97.4	1.5	137850	2 AC133863	AC133863 Medicago
31	97.2	1.5	97146	8 AC007259	AC007259 Arabidops
32	96.4	1.5	4793	8 ATLECRCK	U98010 A.thaliana
33	96.4	1.5	13558	8 ATT93161	Y93161 Arabidops
34	96.4	1.5	91851	8 ATT6122	AL13869 Arabidops
35	96.4	1.5	110000	2 PFMA113_24	Continuation (25 of
36	96.4	1.5	22581	2 BX537105	BX537105 Danio rer
37	96	1.5	114434	2 AC147011	AC147011 Medicago
38	95.6	1.5	106411	8 AP004956	AP004956 Lotus cor
39	95.2	1.5	198743	2 BX530070	BX530070 Danio rer
40	95	1.5	151344	2 AC131343	AC131343 Oryza sat
41	94.6	1.5	82612	8 ATT3185	AL163491 Arabidops
42	93.8	1.5	70382	2 AC147471	AC147471 Medicago
43	93.8	1.5	105246	8 AC139747	AC139747 Medicago
44	93.8	1.5	146415	5 AL929469	AL929469 Zebrarfish
45	93.8	1.5	193499	8 CINS08C94	AL732381 Oryza sat

## ALIGNMENTS

RESULT 1  
LOCUS AF007776 6426 bp DNA linear PLN 21-NOV-1997  
DEFINITION Candida albicans retrotransposon pal, complete sequence.  
ACCESSION AF007776  
VERSION AF007776.1 GI:2636718  
KEYWORDS  
SOURCE  
ORGANISM Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
Saccharomycetales; mitoscopia; Saccharomycetales; Candida.  
REFERENCE  
1 (bases 1 to 6426)  
Matthews,G.D., Goodwin,T.J., Butler,M.I., Berryman,T.A. and  
Poulter,R.T.  
TITLE  
pal, a highly unusual Ty1/copia retrotransposon from the

pathogenic yeast *Candida albicans*  
 JOURNAL J. Bacteriol. 179 (22), 7118-7128 (1997)  
 MEDLINE 98037512  
 PUBMED 9371461  
 REFERENCE 2 (bases 1 to 6426)  
 AUTHORS Matthews,G.D., Goodwin,T.J.D., Butler,M.I., Berryman,T.A. and  
 Poulter,R.T.M.  
 TITLE Direct Submission  
 JOURNAL Submitted (11-JUN-1997) Department of Biochemistry, University of  
 Otago, Box 56, Dunedin, New Zealand

FEATURES  
 source  
 1..6426  
 /organism="Candida albicans"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:5476"  
 /transposon="retrotransposon pCal"  
 1..280  
 281..291  
 /note="minus-strand primer-binding site"  
 398..1372  
 /note="ORF1"  
 /codon\_start=1  
 /transl\_table=12  
 /product="GAG protein"  
 /protein\_id="AAC49879.1"  
 /db\_xref="GI:2636720"  
 /translation="MSSAKNDNEGKVMESYVDQANATSKYDEHAKAPNNLFIKENDL  
 PKLAVNGKSVDKNNEBKRYFHVAVPYVELLDYNNKDKPKYKXVGGYFTGNCLOM  
 CLOSIIFDRFLIMISKPLKLOKEANIKKAYDAVTKSKDYTTYSKLSFVNEHLL  
 VVCYNLPYLSQVEBKLEELIYNTSNVDEYRSLPNIIGVLYFNHVKKEALSLPLN  
 IHASYKRMIOQNDSTVLPSCSTIAEMCDHPYARLVIPSKRYELNIVLPAPE  
 KPKGPEENSESQCKSKSKRKNKHKPDNDKGEKEKEKETSSE"  
 1370..1372  
 /note="potentially suppressed stop codon"  
 <1373..6103  
 /note="ORF2"  
 /codon\_start=1  
 /transl\_table=12  
 /product="POL protein"  
 /protein\_id="AAC49878.1"  
 /db\_xref="GI:2636719"  
 /translation="KTGAASINCVMNHNHNSKTPPYENSHSLNASLNNVNFGRN  
 KYLYVDGATISVNNKDIISNVKDIIEVSADGATLEADICGDLIRIGVISILIE  
 NTLVLPESFNLVSLKQIEERGFNVLITKESVIVNGVAPETIIASKNADYMGQ  
 FSEBLECDPYDGLADMLSNANODDKSKSMENSEYOEHDYSRALINSLEVDL  
 DVEISPYVEBOLPTGQNDIYNFHLMSNEMSEIKILQYQGLVHTKESLOKIA  
 DCKVLLSNKQPSNHSERKASRRHRLHCDLGPFSNNKWVLTSLIDEHTGYI  
 EGIITPKKVDLLIQRLKTNMNPKNKVFYFSDMAPEPPOEDLAEQIMETIAA  
 YSPBLNGAEVYKLLIDQITRIYVTLGPOILKIYVYIOYSTIMNHTRRSLKGT  
 PYGCTYQLSEGNFRFPALDCCVTFSALEKNRYGVSTGAPSSIMGVIGASQC  
 FSYVLLNNRCDIILSPNRLRSYEVINSYKLLSTTPMSHIVPAEIQGQSGA  
 QYEVGTVSEEDYNTNDVMMPKESYVQSPASFTLTGNSNKEYVINDPVOITEN  
 PDDFENPQLTSESHDWSYKSDENPKPSLHETPGDNPVSKPQSGTSTVYGSK  
 EPIIHTKIDAPSIQGRHRSSTAYOGLSHQPTGPASESLSGTDFRGYVARE  
 TVSEDMHSDPYTSARDEQNPISANKRYEKIDGENISIPRGDIAEPIIMNSIV  
 EOSNVEITDAGNPTQDEVSQEKILNEDTIDYTVAKVLENKISIPGINSIDHTELA  
 TDSGNSNSTESDIQSKOEISPVINEKNETITQHLISILADKLDFEFYVNDLEIN  
 VINDIDIANPDPDENNDQMMESFNHMSRPAKKYFEKEVKEKIKTQKESLDT  
 TDPRAIRLVNTGETKRIEKPKEVEPIITVKNLSQKYSYVTSAGTIVNPKYSLA  
 VYNKIDYNDPGMISMAEELKPKSKDVEVEPIPTGVKPSMWHTETDLSLQVY  
 RKSRCVYHGNOKKELIDYPRVSAPVIDLVTILTLTIGELGIMTOHLDVSAAYN  
 ASITSNPTIYFPRKSYVPLKKNHCKMLKRSYVGLKQSGLEHMYIKRYVEDIPQVTL  
 HNDGPHIEBESYIYGLYVDILNQSQKVIDNPOLRDKHEVYKFGESLNTL  
 GIEFKETSEGYILQERFLKLLDFKLDSDYGNLIPINDKEKVAIIREVNPEN  
 DPEKPNETLLDPKAKLYQSGVSLMAATNPDISVYVNSIGSKSAMPNVHDEK  
 LIYCLARYKNSGHIEXYRNLNIPKPSFVIEGSDASAPGDRKISGTLIYVNG  
 NLVQATKQOTVIAQSSAACEMALNTMLKAIIEIKHMLDGEVGIKICHQNDQV  
 IKVLNNYCHHPRPIDCYKFLROLINDKFSISVYNTNVDVNCMKCULSRAPKAF  
 VEGMKRLEIDNQSTONATTAE"  
 1373..1380  
 /note="Region: purine-rich tract"  
 1381..1455  
 /note="Region: potential pseudoknot"

misc\_feature 3455..3465  
 /note="Region: polypurine tract 2"  
 misc\_feature 6136..6146  
 /note="Region: polypurine tract 1"  
 LTR 6147..6426

ORIGIN

Query Match 100.0%; Score 6426; DB 8; Length 6426;  
 Best Local Similarity 100.0%; Freq. No. 0;  
 Matches 6426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TGTGGTGTGCACTATTTGTGTCAGAACTGATCAATGAAATGATGTTATTATGA 60  
 1 TGTGGTGTGCACTATTTGTGTCAGAACTGATCAATGAAATGATGTTATTATGA 60  
 61 GAATGAAAATTTTCCATCAGACATCAGGATGATGAGAACTAATCTATTTGTGTG 120  
 61 GAATGAAAATTTTCCATCAGACATCAGGATGATGAGAACTAATCTATTTGTGTG 120  
 121 ATTAATTAAGGATGATAATATCCACATCCAGATATGACAGATAGAAAGAGAGATT 180  
 121 ATTAATTAAGGATGATAATATCCACATCCAGATATGACAGATAGAAAGAGAGATT 180  
 181 TCAATATATATCTTGTGATATATTAATCTTGTCTATTCATATACACTAGAGCTGT 240  
 181 TCAATATATATCTTGTGATATATTAATCTTGTCTATTCATATACACTAGAGCTGT 240  
 241 AACGCTGATCTCGAGTAAGAAAGTTATATTCACAGATTAGAGAGTGAATGAT 300  
 241 AACGCTGATCTCGAGTAAGAAAGTTATATTCACAGATTAGAGAGTGAATGAT 300  
 301 AATCATTTTCCGTCGCAATTTAGCGTTGATTAATTCAGTCTCAGATTGTAATT 360  
 301 AATCATTTTCCGTCGCAATTTAGCGTTGATTAATTCAGTCTCAGATTGTAATT 360  
 361 GATAGTTTGAAGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420  
 361 GATAGTTTGAAGTTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 420  
 421 TAACGAGGAGAGTCAATGAAAGTGTGATCAAGCTATATCTATTAAGTGAAGTGA 480  
 421 TAACGAGGAGAGTCAATGAAAGTGTGATCAAGCTATATCTATTAAGTGAAGTGA 480  
 481 AATATCAAGCTATATCTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 540  
 481 AATATCAAGCTATATCTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 540  
 541 CGTCGGTATCGAAAAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGA 600  
 541 CGTCGGTATCGAAAAGCGTGAATGAATGAATGAATGAATGAATGAATGAATGA 600  
 601 TTAACCCGATGTTTGAATTTTCTGACTATATCTTAAGATAATTCGAAGTTAA 660  
 601 TTAACCCGATGTTTGAATTTTCTGACTATATCTTAAGATAATTCGAAGTTAA 660  
 661 AAAGGATGAAGTATTTTATTTTACGTTGTTGTTTCAATGTTTACAGTCCATTT 720  
 661 AAAGGATGAAGTATTTTATTTTACGTTGTTGTTTCAATGTTTACAGTCCATTT 720  
 721 TATATGTTCAATGATCATATTTTACGTTGTTGTTTCAATGTTTACAGTCCATTT 780  
 721 TATATGTTCAATGATCATATTTTACGTTGTTGTTTCAATGTTTACAGTCCATTT 780  
 781 CTATATCAAGTCTTATGATGCTGTATCTTAATCTTAAGATTTACATTTACATGA 840  
 781 CTATATCAAGTCTTATGATGCTGTATCTTAATCTTAAGATTTACATTTACATGA 840  
 841 GATCTGCGAGTGTGTAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 900  
 841 GATCTGCGAGTGTGTAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 900  
 901 TTGCTGCGAGTGTGTAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 960  
 901 TTGCTGCGAGTGTGTAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAG 960

Db 901 TTGCTGCAAGGTGGAAGAGAACTTGAGAGAAATCTACACACTTCAACGTTTGCGA 960  
 QY 961 TGAGATATGCGTAGTCTTCCAAATCTAGTCAAGTCTTGACTTCAATCATGTGAA 1020  
 Db 961 TGAGATATGCGTAGTCTTCCAAATCTAGTCAAGTCTTGACTTCAATCATGTGAA 1020  
 QY 1021 GAAATCAAGGCTTAAATTTGTTTGAATATTCATGCTTCATATCTCAAGTGGAT 1080  
 Db 1021 GAAATCAAGGCTTAAATTTGTTTGAATATTCATGCTTCATATCTCAAGTGGAT 1080  
 QY 1081 TCAAGTGAACAATGATACATCATGCTCCAGTGTCTCTACCAATGCTGAAGAAATGTG 1140  
 Db 1081 TCAAGTGAACAATGATACATCATGCTCCAGTGTCTCTACCAATGCTGAAGAAATGTG 1140  
 QY 1141 TGATCATCTGATTTATGCTAGATTGGTTGACATTCGACAAACAATRTGAATCTAATCT 1200  
 Db 1141 TGATCATCTGATTTATGCTAGATTGGTTGACATTCGACAAACAATRTGAATCTAATCT 1200  
 QY 1201 TATGTTAGTTTACAGAGACCGAGAGAAACCAAGAGAGAAACGAGGAGAACTCATCTGGA 1260  
 Db 1201 TATGTTAGTTTACAGAGACCGAGAGAAACCAAGAGAGAAACGAGGAGAACTCATCTGGA 1260  
 QY 1261 ACAATCTCAAAAAGAAAGAACTGAAATCAAGAAAGAAATTAAGAAACATCCAAATGAGA 1320  
 Db 1261 ACAATCTCAAAAAGAAAGAACTGAAATCAAGAAAGAAATTAAGAAACATCCAAATGAGA 1320  
 QY 1321 TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1380  
 Db 1321 TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1380  
 QY 1381 TGCTGCTTCTATTAATGTTGTAATGATATATCATATATGACGAAAAACAGGTTTCCAGT 1440  
 Db 1381 TGCTGCTTCTATTAATGTTGTAATGATATATCATATATGACGAAAAACAGGTTTCCAGT 1440  
 QY 1441 AGAAAAATCTCATCTCTTAATGCTTCTTGAAGTAAATTTAAAGTTTAAAGTT 1500  
 Db 1441 AGAAAAATCTCATCTCTTAATGCTTCTTGAAGTAAATTTAAAGTTTAAAGTT 1500  
 QY 1501 TAAAGATATCTAGTGTATGATGATGCTGTCGACAAATCTGTGTGAACAAATTAAGATAT 1560  
 Db 1501 TAAAGATATCTAGTGTATGATGATGCTGTCGACAAATCTGTGTGAACAAATTAAGATAT 1560  
 QY 1561 ATGCTGATGTTAAGAGAGCAACAAATTAAGTTTCTGTTGCTGATGCTGATTTAGA 1620  
 Db 1561 ATGCTGATGTTAAGAGAGCAACAAATTAAGTTTCTGTTGCTGATGCTGATTTAGA 1620  
 QY 1621 AGCAATATGTTATGTTGATCTTAATATGAGAGTGGTATGTTGCTGATGCTGATTTAGA 1680  
 Db 1621 AGCAATATGTTATGTTGATCTTAATATGAGAGTGGTATGTTGCTGATGCTGATTTAGA 1680  
 QY 1681 TACATTTGATTTTACAGAGAGTTCTTTAATCTGTGAGTTTGAACAAATTTGAAGACG 1740  
 Db 1681 TACATTTGATTTTACAGAGAGTTCTTTAATCTGTGAGTTTGAACAAATTTGAAGACG 1740  
 QY 1741 AGGATTTAATGTTCTTAATTAAGAGATCAAGTATGTTATTAACCAAAATGTGGCTCC 1800  
 Db 1741 AGGATTTAATGTTCTTAATTAAGAGATCAAGTATGTTATTAACCAAAATGTGGCTCC 1800  
 QY 1801 TACATTTATGTTGCTTCAAGAGAAATGCTGCTGATCTTATATGAGGCTCTCAATTCAGTGA 1860  
 Db 1801 TACATTTATGTTGCTTCAAGAGAAATGCTGCTGATCTTATATGAGGCTCTCAATTCAGTGA 1860  
 QY 1861 AGAATCTTTAGATGTATTTGATTTATGATGATGTTGAGATGTTGCTCAATTCAGTGA 1920  
 Db 1861 AGAATCTTTAGATGTATTTGATTTATGATGATGTTGAGATGTTGCTCAATTCAGTGA 1920  
 QY 1921 CCAATATGCAAGATTAATCAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1980  
 Db 1921 CCAATATGCAAGATTAATCAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1980  
 QY 1981 TAGTCTGAGCATTAAATTAATTTCTTTGACGAGGTTGATGTTTGAATGTTGAATTTTC 2040  
 Db 1981 TAGTCTGAGCATTAAATTAATTTCTTTGACGAGGTTGATGTTTGAATGTTGAATTTTC 2040

QY 2041 CCCATATGAGATTTGAACAAATGCTACCAACTGAGATTAAGAACATATTTATTTTCCA 2100  
 Db 2041 CCCATATGAGATTTGAACAAATGCTACCAACTGAGATTAAGAACATATTTATTTTCCA 2100  
 QY 2101 TTGATGTCAAATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
 Db 2101 TTGATGTCAAATTCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2160  
 QY 2161 CGTACTTCACTTCAAAAAGAGATCTTCAAAAAGATTTGCTGATTTGATGATGATGAT 2220  
 Db 2161 CGTACTTCACTTCAAAAAGAGATCTTCAAAAAGATTTGCTGATTTGATGATGATGAT 2220  
 QY 2221 ATGGAATGCAACAGAGAGTCAACATATCATTCAGAAAGAAAAAGCTTCAGAGAGACA 2280  
 Db 2221 ATGGAATGCAACAGAGAGTCAACATATCATTCAGAAAGAAAAAGCTTCAGAGAGACA 2280  
 QY 2281 TGAGAGACTTCATGTTGATGATCTCGGTCCATTTAGTCCGAAAAATTAACAAGTGTATT 2340  
 Db 2281 TGAGAGACTTCATGTTGATGATCTCGGTCCATTTAGTCCGAAAAATTAACAAGTGTATT 2340  
 QY 2341 AACGCTGTTATGATGAAACATACGAGTTACATTTGAAGAAATTTACTTAAGACAGAAA 2400  
 Db 2341 AACGCTGTTATGATGAAACATACGAGTTACATTTGAAGAAATTTACTTAAGACAGAAA 2400  
 QY 2401 GGTAAAGSACTCTTAAATCAACGATTAAGATTTGAAGATTTGCGTTTAAAGATTAAGT 2460  
 Db 2401 GGTAAAGSACTCTTAAATCAACGATTAAGATTTGAAGATTTGCGTTTAAAGATTAAGT 2460  
 QY 2461 GGCATACCTTCAAGAGATTAAGTCTCGAGTTCCACAACTTCTGATTTAGTGAAT 2520  
 Db 2461 GGCATACCTTCAAGAGATTAAGTCTCGAGTTCCACAACTTCTGATTTAGTGAAT 2520  
 QY 2521 CGGTATTTGAGAGGACATTAAGCGGCAATCTGCTGAGCTTAATGCTGCGGAGGT 2580  
 Db 2521 CGGTATTTGAGAGGACATTAAGCGGCAATCTGCTGAGCTTAATGCTGCGGAGGT 2580  
 QY 2581 TGTAAATTAATGATTTTCAACAGATTTACAGATTCGTTGACACATTTGATGATGAT 2640  
 Db 2581 TGTAAATTAATGATTTTCAACAGATTTACAGATTCGTTGACACATTTGATGATGAT 2640  
 QY 2641 ACTCAAGTATTTATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2700  
 Db 2641 ACTCAAGTATTTATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2700  
 QY 2701 TCGTTACTCAAGGAGCAAAACCCCTTATGATGATGATGATGATGATGATGATGATGAT 2760  
 Db 2701 TCGTTACTCAAGGAGCAAAACCCCTTATGATGATGATGATGATGATGATGATGATGAT 2760  
 QY 2761 CTACCGGTTTCTTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2820  
 Db 2761 CTACCGGTTTCTTTTGCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2820  
 QY 2821 CGGTAGGAGTTTACCACTTAAGAGAGCTCTTTCATGATGATGATGATGATGATGATGAT 2880  
 Db 2821 CGGTAGGAGTTTACCACTTAAGAGAGCTCTTTCATGATGATGATGATGATGATGATGAT 2880  
 QY 2881 CTACGCTAGAGATTTGTTAGTATTAAGTGTGCTTAAAAAAATATGCGGTGATATTAAT 2940  
 Db 2881 CTACGCTAGAGATTTGTTAGTATTAAGTGTGCTTAAAAAAATATGCGGTGATATTAAT 2940  
 QY 2941 CCTTACGCTTAAATGCTGATTAATGGAAGCTATGATGATGATGATGATGATGATGATGAT 3000  
 Db 2941 CCTTACGCTTAAATGCTGATTAATGGAAGCTATGATGATGATGATGATGATGATGATGAT 3000  
 QY 3001 CTATACCACTACCACTATGATCAACATTTGCTTCAATGCTGATGATGATGATGATGATGAT 3060  
 Db 3001 CTATACCACTACCACTATGATCAACATTTGCTTCAATGCTGATGATGATGATGATGATGAT 3060  
 QY 3061 ACTGGGCGCTCAGTACGAGATGACGAGAAATATGATGATGATGATGATGATGATGATGAT 3120  
 Db 3061 ACTGGGCGCTCAGTACGAGATGACGAGAAATATGATGATGATGATGATGATGATGATGAT 3120

QY 3121 TGACGTGATGCACATGCCCAAGAGTCAATTCAGTTCAGCCAGCATCGTTTACTTTAAC 3180  
Db 3121 TGACGTGATGCACATGCCCAAGAGTCAATTCAGTTCAGCCAGCATCGTTTACTTTAAC 3180  
QY 3181 TAGCGGTACAGTCTCAACGAATATGTTATTAATGATGATCCAGTACAGATTACCATTTGA 3240  
Db 3181 TAGCGGTACAGTCTCAACGAATATGTTATTAATGATGATCCAGTACAGATTACCATTTGA 3240  
QY 3241 GAATCCCGATGATTTTCTTAAACCTCTTCAAATTAAGTGAATACAGATATGTTATC 3300  
Db 3241 GAATCCCGATGATTTTCTTAAACCTCTTCAAATTAAGTGAATACAGATATGTTATC 3300  
QY 3301 CGAAGTAAATCGATGAGATCTTAAACCGATCTCAAGATCCAGATTAACCTGGGATTA 3360  
Db 3301 CGAAGTAAATCGATGAGATCTTAAACCGATCTCAAGATCCAGATTAACCTGGGATTA 3360  
QY 3361 TCCGGTGTCTAAACCTCTCTCAACTGGTACCGAGACTTCAATTAAGGAACTTAAAGA 3420  
Db 3361 TCCGGTGTCTAAACCTCTCTCAACTGGTACCGAGACTTCAATTAAGGAACTTAAAGA 3420  
QY 3421 GCCTATTTCAC 3480  
Db 3421 GCCTATTTCAC 3480  
QY 3481 GGAATCTACTGCTCAGGTGAGTATCAACACACACACACACACACACACACACACAC 3540  
Db 3481 GGAATCTACTGCTCAGGTGAGTATCAACACACACACACACACACACACACACACAC 3540  
QY 3541 GGAATCTACTGCTCAGGTGAGTATCAACACACACACACACACACACACACACACAC 3600  
Db 3541 GGAATCTACTGCTCAGGTGAGTATCAACACACACACACACACACACACACACACAC 3600  
QY 3601 AGAAGATTGGATCTTCTGACTACCCGAACTAGTCTGAAGATGAACAGCAAAATCC 3660  
Db 3601 AGAAGATTGGATCTTCTGACTACCCGAACTAGTCTGAAGATGAACAGCAAAATCC 3660  
QY 3661 CTCGTTACTGCTAATAAGATCGGGTACCTGAAAAATGATGAGGAGAAATATTTC 3720  
Db 3661 CTCGTTACTGCTAATAAGATCGGGTACCTGAAAAATGATGAGGAGAAATATTTC 3720  
QY 3721 ATTTCGGGGGGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780  
Db 3721 ATTTCGGGGGGGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3780  
QY 3781 TGAACACAGAGATGCTGTGTAACGTCCAAATCAAGACGAGTTCTCAAGGAGAAAT 3840  
Db 3781 TGAACACAGAGATGCTGTGTAACGTCCAAATCAAGACGAGTTCTCAAGGAGAAAT 3840  
QY 3841 ACTTATGACAACTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 3900  
Db 3841 ACTTATGACAACTGATATGATGATGATGATGATGATGATGATGATGATGATGAT 3900  
QY 3901 CTCCTCTATTATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Db 3901 CTCCTCTATTATTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
QY 3961 CAATTCACACAGATTCGACATTCATTCGAAAAATGAAATCACCACTGATTAATGAA 4020  
Db 3961 CAATTCACACAGATTCGACATTCATTCGAAAAATGAAATCACCACTGATTAATGAA 4020  
QY 4021 AAATCTGAAATTAATCCAAAAACATGAAAGTATCTTCTGATTAAGATGAGATG 4080  
Db 4021 AAATCTGAAATTAATCCAAAAACATGAAAGTATCTTCTGATTAAGATGAGATG 4080  
QY 4081 ATTGAAACGATTAATGTTGATGAATGAGATGATGATGATGATGATGATGATGATG 4140  
Db 4081 ATTGAAACGATTAATGTTGATGAATGAGATGATGATGATGATGATGATGATGATG 4140  
QY 4141 AGCTATCCACATTCACAGATGAAATTAATGATGATGATGATGATGATGATGATGAT 4200  
Db 4141 AGCTATCCACATTCACAGATGAAATTAATGATGATGATGATGATGATGATGATGAT 4200  
QY 4201 TCATGACATGTCACAGCAAAAGAAATTAACAATTGAGAAAGATTTAAGAAAAAT 4260

Db 4201 TCATGACATGTCACAGCAAAAGAAATTAACAATTGAGAAAGATTTAAGAAAAAT 4260  
QY 4261 TECTGTGCTAAACATTCATTTGATTAACATGATTCAGAGACAAATCAGAGTTTAA 4320  
Db 4261 TECTGTGCTAAACATTCATTTGATTAACATGATTCAGAGACAAATCAGAGTTTAA 4320  
QY 4321 TACTGTTAAACCAAGATGAAATCCCAAGAAAGAGAGTGCCTATCACTGGAAT 4380  
Db 4321 TACTGTTAAACCAAGATGAAATCCCAAGAAAGAGAGTGCCTATCACTGGAAT 4380  
QY 4381 AAACAAAGATTCGCAATCAAGTCAACATGTTTAAAGAAAGTGAACGGTTATTA 4440  
Db 4381 AAACAAAGATTCGCAATCAAGTCAACATGTTTAAAGAAAGTGAACGGTTATTA 4440  
QY 4441 CCCCAAGAGATTTTACATGCGGTGCTAAACAAATGAGATTAATGATCCGGATGAT 4500  
Db 4441 CCCCAAGAGATTTTACATGCGGTGCTAAACAAATGAGATTAATGATCCGGATGAT 4500  
QY 4501 AAAGTCAATGATGCTGAACCTGAAATGAAATTTAGATCAAAAGATTTTACGAAGATTC 4560  
Db 4501 AAAGTCAATGATGCTGAACCTGAAATGAAATTTAGATCAAAAGATTTTACGAAGATTC 4560  
QY 4561 AATTCACACCGGTGGAACCTTATTCATGCGGTGCGTATGATGATGATGATGATGAT 4620  
Db 4561 AATTCACACCGGTGGAACCTTATTCATGCGGTGCGTATGATGATGATGATGATGAT 4620  
QY 4621 TCTCAAGATGTTTGTGGAATCACTGTTGTTGTCATGAGCAACACAAAGAA 4680  
Db 4621 TCTCAAGATGTTTGTGGAATCACTGTTGTTGTCATGAGCAACACAAAGAA 4680  
QY 4681 ATTGATTAATGACCTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
Db 4681 ATTGATTAATGACCTTTTATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
QY 4741 ATTGACAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800  
Db 4741 ATTGACAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800  
QY 4801 GTATCTAAATGCTCTTATTAATCAATTCATTCATTCATTCATTCATTCATTCATTC 4860  
Db 4801 GTATCTAAATGCTCTTATTAATCAATTCATTCATTCATTCATTCATTCATTCATTC 4860  
QY 4861 ACCTTTGAAGAAACCAATGTTGTTTGAACGTTCTGATGAGGTTAAACAGTC 4920  
Db 4861 ACCTTTGAAGAAACCAATGTTGTTTGAACGTTCTGATGAGGTTAAACAGTC 4920  
QY 4921 GGGTTGGAATGATGATCACTATCAAAAGATTTGGAAGACATTTGTTTACTCAAGT 4980  
Db 4921 GGGTTGGAATGATGATCACTATCAAAAGATTTGGAAGACATTTGTTTACTCAAGT 4980  
QY 4981 TTTTCAACATGATGTTTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 5040  
Db 4981 TTTTCAACATGATGTTTATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 5040  
QY 5041 TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5100  
Db 5041 TTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5100  
QY 5101 GGATCAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5160  
Db 5101 GGATCAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5160  
QY 5161 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5220  
Db 5161 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5220  
QY 5221 ATTACTTAAGATTTCAACTGATGATGATGATGATGATGATGATGATGATGATGATG 5280  
Db 5221 ATTACTTAAGATTTCAACTGATGATGATGATGATGATGATGATGATGATGATGATG 5280  
QY 5281 TGACAAATTAAGAGTTGCAATTAATCGTGAAGAACTTATTCAGGAATGATTTGA 5340

Db 5281 TGACAAATATGAAAAAGTTGCAATTAATTCGTAAGAAAAAGTTATTCACAGAAATGATTTTGA 5340  
 Qy 5341 AAAGGTTCCGATGAGACATTCGTTGACCCCTGATGCTAAAAAATATACCAAAAGTGGTGT 5400  
 Db 5341 AAAAGTTCCGATGAGACATTCGTTGACCCCTGATGCTAAAAAATATACCAAAAGTGGTGT 5400  
 Qy 5401 TGGCTGCTTTTATGAGGCTGCGACAAACAGCTCCAGATATATCGCTGATGAAATTC 5460  
 Db 5401 TGGCTGCTTTTATGAGGCTGCGACAAACAGCTCCAGATATATCGCTGATGAAATTC 5460  
 Qy 5461 GTTGGGCTCTAAATGCAAAATCCAAATGTCATGATATAGAAATGATTTATGCT 5520  
 Db 5461 GTTGGGCTCTAAATGCAAAATCCAAATGTCATGATATAGAAATGATTTATGCT 5520  
 Qy 5521 TAGATATATCAAAAAATAGCATGGGATATACATTTAGTACAAAAAGAAACAGATTGAATAT 5580  
 Db 5521 TAGATATATCAAAAAATAGCATGGGATATACATTTAGTACAAAAAGAAACAGATTGAATAT 5580  
 Qy 5581 ACCACAAATATCTTTTATGATTCGAAATGTTGAGATCCGTCATTTGACACAGATTTGA 5640  
 Db 5581 ACCACAAATATCTTTTATGATTCGAAATGTTGAGATCCGTCATTTGACACAGATTTGA 5640  
 Qy 5641 ACCACAAATATCTTTTATGATTCGAAATGTTGAGATCCGTCATTTGACACAGATTTGA 5640  
 Db 5641 ACCACAAATATCTTTTATGATTCGAAATGTTGAGATCCGTCATTTGACACAGATTTGA 5640  
 Qy 5641 TAGAAAAATCTATATGATGAACTTTGATTTATGTAATGGAATTTGGTCAATGGGCGAC 5700  
 Db 5641 TAGAAAAATCTATATGATGAACTTTGATTTATGTAATGGAATTTGGTCAATGGGCGAC 5700  
 Qy 5701 CAAAAAACAAGGTCATACGCAAAAGCTCAGACGCTTGTGAATTTGGCTCTTAATTA 5760  
 Db 5701 CAAAAAACAAGGTCATACGCAAAAGCTCAGACGCTTGTGAATTTGGCTCTTAATTA 5760  
 Qy 5761 TACAAATGTTGAAGGCTATGCAAAATCAATTAATGATTTGGGTTTGAAGTAGG 5820  
 Db 5761 TACAAATGTTGAAGGCTATGCAAAATCAATTAATGATTTGGGTTTGAAGTAGG 5820  
 Qy 5821 TAAAGTATCTGTCATCAAGACACCAAGCTGTGATTAAGTTTGAAGAAATCAATTTG 5880  
 Db 5821 TAAAGTATCTGTCATCAAGACACCAAGCTGTGATTAAGTTTGAAGAAATCAATTTG 5880  
 Qy 5881 TCACCAATGTCAGCAATGATATCTGCTATGATTTGTCAGCAATTTGATCAATGATA 5940  
 Db 5881 TCACCAATGTCAGCAATGATATCTGCTATGATTTGTCAGCAATTTGATCAATGATA 5940  
 Qy 5941 AGTATTTTCAATATCTATGTCAGCAATGATATTAATGACCGCATTTGATGACTAGTG 6000  
 Db 5941 AGTATTTTCAATATCTATGTCAGCAATGATATTAATGACCGCATTTGATGACTAGTG 6000  
 Qy 6001 TTTAAGTGTGCTAATTCAAAGCATTGCTGAGGGATGATTAACCGTTAGACTAGA 6060  
 Db 6001 TTTAAGTGTGCTAATTCAAAGCATTGCTGAGGGATGATTAACCGTTAGACTAGA 6060  
 Qy 6061 AGATATATCAAACTGATACAAATGCAATTAACGCGCAATTAAGTGATTTATCATTA 6120  
 Db 6061 AGATATATCAAACTGATACAAATGCAATTAACGCGCAATTAAGTGATTTATCATTA 6120  
 Qy 6121 ATTATGTAATGCTCAATCAGGGAGTGTGGTTTGTGCACATTTTGTGCAGAACTG 6180  
 Db 6121 ATTATGTAATGCTCAATCAGGGAGTGTGGTTTGTGCACATTTTGTGCAGAACTG 6180  
 Qy 6181 ATCAATGAAATGATGCTTATATGAAATGAAATTTTTCATCAACATCAGGTAT 6240  
 Db 6181 ATCAATGAAATGATGCTTATATGAAATGAAATTTTTCATCAACATCAGGTAT 6240  
 Qy 6241 GACAGAACTAAATCTATTTGTAATTAATTAAGGATTAAGAAATCAACATCCAGAA 6300  
 Db 6241 GACAGAACTAAATCTATTTGTAATTAATTAAGGATTAAGAAATCAACATCCAGAA 6300  
 Qy 6301 TTTCAACGAGATAGAGGAGAGGATTCATTAATTAATCTGGAATATACTGCTTCT 6360  
 Db 6301 TTTCAACGAGATAGAGGAGAGGATTCATTAATTAATCTGGAATATACTGCTTCT 6360  
 Qy 6361 AATTCATATACAACTAGACGTGTACACGCTCAATCTCAGGTAAAGAAAGTTATAT 6420  
 Db 6361 AATTCATATACAACTAGACGTGTACACGCTCAATCTCAGGTAAAGAAAGTTATAT 6420

Qy 6421 CCATCA 6426  
 Db 6421 CCATCA 6426  
 RESULT 2  
 AF050215  
 LOCUS  
 DEFINITION  
 Candida albicans Tca2 retrotransposon gag polyprotein (gag) and pol  
 polyprotein (pol) genes, complete cds.  
 ACCESSION  
 AF050215  
 VERSION  
 AF050215.1 GI:3273716  
 KEYWORDS  
 ORGANISM  
 Candida albicans  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 REFERENCE  
 1 (bases 1 to 6980)  
 Goodwin,T.J.D. and Poulter,R.T.M.  
 Temperature- and strain-dependent expression of the pcal  
 retrotransposon of Candida albicans  
 JOURNAL  
 Unpublished  
 2 (bases 1 to 6980)  
 Goodwin,T.J.D. and Poulter,R.T.M.  
 Direct Submission  
 Submitted (20-FEB-1998) Department of Biochemistry, University of  
 Otago, Cumberland Street, Dunedin, New Zealand  
 Location/Qualifiers  
 FEATURES  
 source  
 1..6980  
 /organism="Candida albicans"  
 /mol\_type="genomic DNA"  
 /strain="H06759"  
 /db\_xref="taxon:5476"  
 /chromosome="3"  
 /map="between CDRI and DVN1"  
 202..206  
 /note="5' target site duplication"  
 207..6632  
 /rpt\_family="Tca2 retrotransposon"  
 /rpt\_type="dispersed"  
 207..486  
 /note="5' long terminal repeat"  
 487..518  
 /note="complements tRNA Arg (UCU) fragment"  
 604..1578  
 /gene="gag"  
 604..1578  
 /gene="gag"  
 /codon\_start=1  
 /transl\_table=12  
 /product="gag polyprotein"  
 /protein\_id="AAC24820.1"  
 /db\_xref="GI:3273717"  
 /translation="MSKAKDNDNEKVESVQANAKISKVDHIKAREMFLPIKENDL  
 PKLAVGNQKSVKRNBERKYPHVAIPDVLEFLDNDPKRKVKKEGITYFGWCLQIOM  
 CLOSIPDRRLIMISKLPHLQKENALIKAAVDATKSKDYITTSKLSFVNEHEL  
 WVCYNLPYLSOVEKLEELIYNISNVDEYVSLPNLIGQVYFNHVKSEALSLFLN  
 IHSYYSKNIQADNDTSVLPSCSTIAEEMCDHPYARLVDIPSNKYELMLIVSLPAPE  
 KPKGKPENSSEGSQKNSKRNRKHPKSDNDKGEKEKEKTSSE"  
 1576..1578  
 /gene="gag"  
 /note="potentially suppressed stop codon"  
 <1579..6309  
 /gene="pol"  
 <1579..6309  
 /gene="pol"  
 /note="includes protease, integrase, reverse  
 transcriptase, RNase H"  
 /codon\_start=1  
 /transl\_table=12  
 /product="pol polyprotein"  
 /protein\_id="AAC24821.1"

		/db_xref="GI:3273718"	
		/translation="KTGAASINCVMIHNCSTTPPEVNSHLSNLSNLTNPNKXGLRPN	
		KYLIVDTGATISVNNKDIISNVADATIEVSVDGATLEDCIGDILIRGAVISTILE	
		NTLYIPSSFNVLISLQIERGFVLLITKESVIVFNQNVAPTIISARKNAALYMPQ	
		FSESLCEDPYDLADLNDNANDKDSNMENSEYCHDSSALINSLIEVYL	
		DVEISPYGECLDPTGDKNDIVNFHLSNMASIEKILQKYGGLVYHTEKSELQIA	
		DCKYCLISNAKORSHNHSERKASRHRERHLCDTIGPERENKWTYLIVDTHCYI	
		EGITTKRKYKVDLIDRLKIWNRFNRYVPSDNPPEPQSDLAERIMWETIAA	
		EYELNGLAEVNLILQIIRLYVTLGPIILKLYIVYISITMYSTINHPRLKQOT	
		PYGYVQLSENGFRPPFAIDCVTFENALENKRYGVTSKGPASSIMGIVASYDC	
		PSYVLLKMKRCDILSPNVAILRSYVINSYLNLSITTMSEIIVMAEGLQORQGA	
		QYEVGYVESEEDNTNDVMHMKESYVQPAFTLTGNSNEVYINDPVQITLEN	
		PDPSNPLQLTSESDMVSEKSDENKPSLHELPGDNVSKPQIGTSTVIGSX	
		EPITNTHKDAISIOGRDHRKSESTAVGLSHOPTGPASBEKSLGDTMHEGVNKE	
		TVSEDMHTSDPYETSADDEOONPSLSNKKRYTEKIDEGENISPRGGDDSVYVNE	
		EQSVNVEDHGENSPIDQESOSGRILNEQDIDVTAKVLENKISPIINSLDHTELA	
		TVDSNDNSNTESDIQSKNELSPVINEKNTBIIQKHIESIADRLKEPFIYNDLEIN	
		VINDDDIAENPLPDENNDDVMESPDNNHSMGRKKCYFEKENVKELAGTSHLDT	
		TDPREAIRVNTGETKRIERKREVPITVILNKRSQYKSYVRSQGTIVNPKRYHA	
		LVNKIDYNDPMGIMSKMAELEKPSKQVYEEVPIPTGVKDISMGWTEKIDELKGV	
		RKSRCVHGNROKEXKLDYDPVSVPIDILVTIRLILIGCELMWITIOHLDVSAVLN	
		ASTHNSPIYVFPKSVPLKKNHMLKRSVGLKQSGPEWHTIRKVLIEDIGSAYLN	
		HNDGLFHI EYBESVLYGLVDDILMWSSQKVIDFVQLDKEHVKVKGTSVYL	
		GIERFKTESGYILISQEKFLKRLKDFELDSDYGNLIPWINDXKEYVALIRENVNEN	
		DPEKVPNETLIDPDARKLYOSGVGSLMAANTRPDISVNVSLGSKSANPNVDYEX	
		LICYLRIKSMGYHIEYKANRLNIPKSEVIGCFSPASAPGLDERSISGTLIYNG	
		NLVOMATRKQTVIWOSSACEMALANTMTLKAIEIKNMLDLGFEVKGKIHCHODNCAV	
		IKUTRNHYCHPHRPIDI CYKELRQILINDKXFSISYVNTNNVADMTKCLSRKXFAF	
		VEGATKRLDIEDNOTSIQNAITAE"	
misc_feature	1579..1586	/gene="pol"	
		/note="purine-rich tract"	
misc_feature	1587..1661	/gene="pol"	
		/note="encodes putative pseudoknot"	
misc_feature	3661..3671	/gene="pol"	
		/note="polyurine tract 2"	
misc_feature	6342..6352	/note="polyurine tract 1"	
LTR	6353..6632	/note="3' long terminal repeat"	
misc_feature	6633..6637	/note="3' target site duplication"	
ORIGIN			
Query Match	99.9%; Score 6421.2; DB 8; Length 6980;		
Best Local Similarity	100.0%; Pred. No. 0;		
Matches 6423; Conservative	0; Mismatches 3; Indels 0; Gaps 0;		
QY	1	TGTTGGTTTGCACTATTTTGTGTCAGAACCTGATCAATGAAGAAATGATGTTATTTGA	60
DB	207	TGTTGGTTTGCACTATTTTGTGTCAGAACCTGATCAATGAAGAAATGATGTTATTTGA	266
QY	61	GAATGGAATTTTTCATCAACATCAGTGATGACAGAACTAAATATATTGTAGT	120
DB	267	GAATGGAATTTTTCATCAACATCAGTGATGACAGAACTAAATATATTGTAGT	326
QY	121	ATAAATGAAGGTATGAATTAACCAATCCAGAAATATCAAGAGATNAGAGGAGAGTT	180
DB	327	ATAAATGAAGGTATGAATTAACCAATCCAGAAATATCAAGAGATNAGAGGAGAGTT	386
QY	181	TCAATATATCTCTGTGTAATAATCTGTTCAATTCACATATACCAACTAGAGGT	240
DB	387	TCAATATATCTCTGTGTAATAATCTGTTCAATTCACATATACCAACTAGAGGT	446
QY	241	ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTAGAAGTCGATAGTAT	300
DB	447	ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTAGAAGTCGATAGTAT	506
QY	301	AATCATTTTGGTCCCAATTAAGCTTTGTAATAATCAGTCTCAGATTTGATTAATTT	360
DB	507	AATCATTTTGGTCCCAATTAAGCTTTGTAATAATCAGTCTCAGATTTGATTAATTT	566
QY	361	CATATGTTTGAAGTTGAAGGTACAGAAATTTCAAGATGAGTCCGCAAGAAATGATGA	420
DB	567	GATTAATTTGGAAGTTGAAGGTACAGAAATTTCAAGATGAGTCCGCAAGAAATGATGA	626
QY	421	TACCAAGGAAGGTACATGAGAAAGTGTGATCAAGCTAATGCTATTAAGTGAAGTGA	480
DB	627	TACCAAGGAAGGTACATGAGAAAGTGTGATCAAGCTAATGCTATTAAGTGAAGTGA	686
QY	481	ACATATCAAGGTATATTCATATGCTTTTTCATATAAATTTAATGACTTAACTAGTGC	540
DB	687	ACATATCAAGGTATATTCATATGCTTTTTCATATAAATTTAATGACTTAACTAGTGC	746
QY	541	CGTGGTATCAAGAAACGCTGATTAATGAAATGAAGAAATTTAATTTCCAGTTC	600
DB	747	CGTGGTATCAAGAAACGCTGATTAATGAAATGAAGAAATTTAATTTCCAGTTC	806
QY	601	TTACCCCATGTTTGGAAATTTTGGCTTGACTATATATCTTAAAGATAATTCAGGTTAA	660
DB	807	TTACCCCATGTTTGGAAATTTTGGCTTGACTATATATCTTAAAGATAATTCAGGTTAA	866
QY	661	AAAGGTAGAGGTATTTATTTTAACTGGTGTGTTACAAATGCTTACAGTCATTTT	720
DB	867	AAAGGTAGAGGTATTTATTTTAACTGGTGTGTTACAAATGCTTACAGTCATTTT	926
QY	721	TGATAGTTCAGATGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGAGCAAA	780
DB	927	TGATAGTTCAGATGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGAGCAAA	986
QY	781	CTTAATCAAAAGCTCTTATGATGCTGCTACGAAATCTTAATATATATCCATTACATGA	840
DB	987	CTTAATCAAAAGCTCTTATGATGCTGCTACGAAATCTTAATATATATCCATTACATGA	1046
QY	841	GATCTTGCTGAAGTTTGTAAAGTGTGAACATGAGTATGCTGTTGCTATTAACCTTCATA	900
DB	1047	GATCTTGCTGAAGTTTGTAAAGTGTGAACATGAGTATGCTGTTGCTATTAACCTTCATA	1106
QY	901	TTTGGCTGAAGTGAAGAAACCTTGAGAAATCTCTACAACTTCAACAGTTCGCA	960
DB	1107	TTTGGCTGAAGTGAAGAAACCTTGAGAAATCTCTACAACTTCAACAGTTCGCA	1166
QY	961	TGAGTATGCTGATCTTCCAAATCTCATAGTCAAGTCTTGTGATTTCAATCATGTGAA	1020
DB	1167	TGAGTATGCTGATCTTCCAAATCTCATAGTCAAGTCTTGTGATTTCAATCATGTGAA	1226
QY	1021	GAAATCAAGGCTTTAAGTTGTTTGTGAATTTCAATGCTCTCATATCTCAAAAGTGAT	1080
DB	1227	GAAATCAAGGCTTTAAGTTGTTTGTGAATTTCAATGCTCTCATATCTCAAAAGTGAT	1286
QY	1081	TCAAGCTGACATGATATCATCATGATCTCCAAAGTGTGCTCACTCACTGAGTGAAGATGTG	1140
DB	1287	TCAAGCTGACATGATATCATCATGATCTCCAAAGTGTGCTCACTCACTGAGTGAAGATGTG	1346
QY	1141	TGATCATCTGATTTATGCTAAGTTGTTGACATTTCCAAAGCAAAATATGAACTTAATCT	1200
DB	1347	TGATCATCTGATTTATGCTAAGTTGTTGACATTTCCAAAGCAAAATATGAACTTAATCT	1406
QY	1201	TATTTGTTAGTTTACAGACACAGAGAAACCAAAAGGAAACCAAGAGGAACTCATGGA	1260
DB	1407	TATTTGTTAGTTTACAGACACAGAGAAACCAAAAGGAAACCAAGAGGAACTCATGGA	1466
QY	1261	ACAATCTCAAAAGAAAGAACTGAATCAAGAAAGAAATTAAGAAACATCCCAAAATGAGA	1320
DB	1467	ACAATCTCAAAAGAAAGAACTGAATCAAGAAAGAAATTAAGAAACATCCCAAAATGAGA	1526
QY	1321	TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGG	1380
DB	1527	TAAAGTAAAGGTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGG	1586
QY	1381	TGCTGCTCTATTAATTTGTTGATGATATATCATATATGTCAGCAAAACCAAGTTCAGT	1440
DB	1587	TGCTGCTCTATTAATTTGTTGATGATATATCATATATGTCAGCAAAACCAAGTTCAGT	1646



QY 1441 AGAATTCCTGATCTCTTAATGCTCTTGAAGCTAATGAATTTAAAGGTTAAGGTT 1500  
DB 1647 AGAATATCTCTATCTCTTAATGCTCTTGAAGCTAATGAATTTAAAGGTTAAGGTT 1706  
QY 1501 TAAACAGATCTAGTGTATGATCTGCTGCAATATCTGTGTGAACTAAAGATAT 1560  
DB 1707 TAAACAGATCTAGTGTATGATCTGCTGCAATATCTGTGTGAACTAAAGATAT 1766  
QY 1561 ATTGCTGATGTTAAGAGCAAGCAATGAATTTCTGTGTGATGCTGATTAAGA 1620  
DB 1767 ATTGCTGATGTTAAGAGCAAGCAATGAATTTCTGTGTGATGCTGATTAAGA 1826  
QY 1621 AGCAGATGTTATGTTGATCTTAATTAAGAGTGGTATTTGCTGATTAAGAGAA 1680  
DB 1827 AGCAGATGTTATGTTGATCTTAATTAAGAGTGGTATTTGCTGATTAAGAGAA 1886  
QY 1681 TACATTTGATTTACAGAAAGTCTCTTAATCTGTGATTTGAAACAAATTTGAAGAG 1740  
DB 1887 TACATTTGATTTACAGAAAGTCTCTTAATCTGTGATTTGAAACAAATTTGAAGAG 1946  
QY 1741 AGGATTTAATGTTCTTATTAATAAGAAATCAGTGTATTTAAACCAAAATGCTGCTC 1800  
DB 1947 AGGATTTAATGTTCTTATTAATAAGAAATCAGTGTATTTAAACCAAAATGCTGCTC 2006  
QY 1801 TACTATTTGCTTGAAGAAAGATGCTGATCTCTTAATGAGGTCCTCAATTCAGGA 1860  
DB 2007 TACTATTTGCTTGAAGAAAGATGCTGATCTCTTAATGAGGTCCTCAATTCAGGA 2066  
QY 1861 AGAATCTTAGAATGATTTGATTTGATTAAGAGTGTGAGAAATGTTGCCAATGCTAA 1920  
DB 2067 AGAATCTTAGAATGATTTGATTTGATTAAGAGTGTGAGAAATGTTGCCAATGCTAA 2126  
QY 1921 CCAAGATGCAAAAGTAAATCAAGTATGATGAATGCAAGATTCAGAAACATGATTA 1980  
DB 2127 CCAAGATGCAAAAGTAAATCAAGTATGATGAATGCAAGATTCAGAAACATGATTA 2186  
QY 1981 TAGTCTGAGAGTAAATTAATTTCTTTGACGAGGTTGATTTGATTTGATTTGATTTG 2040  
DB 2187 TAGTCTGAGAGTAAATTAATTTCTTTGACGAGGTTGATTTGATTTGATTTGATTTG 2246  
QY 2041 CCCATATGAGTGAAGCAATTTGCTCAACTGAGAGATGAAGACGATTTTAATTTCCA 2100  
DB 2247 CCCATATGAGTGAAGCAATTTGCTCAACTGAGAGATGAAGACGATTTTAATTTCCA 2306  
QY 2101 TTTGATGTCAAATCATATGTCATTTGAGAAATCTGTGTTGCAAAATTCAGAGGCT 2160  
DB 2307 TTTGATGTCAAATCATATGTCATTTGAGAAATCTGTGTTGCAAAATTCAGAGGCT 2366  
QY 2161 CGTACTTCACTCACTCAAAAGAGAGTCTTCAAAAGATGCTGATTTGATTTGATTTG 2220  
DB 2367 CGTACTTCACTCACTCAAAAGAGAGTCTTCAAAAGATGCTGATTTGATTTGATTTG 2426  
QY 2221 ATGCAATGCAAAAGCAAGATGCAATCATATTTCAAAAGAAAGCTGCAAGAGCA 2280  
DB 2427 ATGCAATGCAAAAGCAAGATGCAATCATATTTCAAAAGAAAGCTGCAAGAGCA 2486  
QY 2281 TGAGAGACTTCACTGATTAATCTCGTCCATTTGATTTGATTTGATTTGATTTG 2340  
DB 2487 TGAGAGACTTCACTGATTAATCTCGTCCATTTGATTTGATTTGATTTGATTTG 2546  
QY 2341 AAGCTGCTTATTAATGAACATACGAGTTACATTTGAAGAAATTTATTAAGACAGAA 2400  
DB 2547 AAGCTGCTTATTAATGAACATACGAGTTACATTTGAAGAAATTTATTAAGACAGAA 2606  
QY 2401 GGTAAAGATCTCTTAATCAACGATTAAGATCTGAAATACGCTTAACGATTAAGAT 2460  
DB 2607 GGTAAAGATCTCTTAATCAACGATTAAGATCTGAAATACGCTTAACGATTAAGAT 2666  
QY 2461 GGCATATCTCAGAAAGTATGATCTCTGAGTTCCACAACTCTGATTTAGCTGAGTT 2520  
DB 2667 GGCATATCTCAGAAAGTATGATCTCTGAGTTCCACAACTCTGATTTAGCTGAGTT 2726  
QY 2521 CGGTATTTGAGAGAGATATAGCGGCAATCTGCTGATTAATGATTTGCTGCGGAGT 2580

DB 2727 CGGTATTTGAGAGAGATATAGCGGCAATCTGCTGATTAATGATTTGCTGCGGAGT 2786  
QY 2581 TGTAAATTAATGATTTTCAACAGATTTACAGAGATCTGTGACACTGTCACAAAT 2640  
DB 2787 TGTAAATTAATGATTTTCAACAGATTTACAGAGATCTGTGACACTGTCACAAAT 2846  
QY 2641 ACTCAAGTATTAATTAATGATTTCAATATTTCAATTAATGATTAACCAACTCCAG 2700  
DB 2847 ACTCAAGTATTAATTAATGATTTCAATATTTCAATTAATGATTAACCAACTCCAG 2906  
QY 2701 TCGTTACCAAGGCAAAACCCCTTAATGATTTGATTTCAATTAATGAGAGAAATTT 2760  
DB 2907 TCGTTACCAAGGCAAAACCCCTTAATGATTTGATTTCAATTAATGAGAGAAATTT 2966  
QY 2761 CTACCGGTTCCCTTTGCCATGATTTGCTGATTTCAATTAATGATTAACCAACTCCAG 2820  
DB 2967 CTACCGGTTCCCTTTGCCATGATTTGCTGATTTCAATTAATGATTAACCAACTCCAG 3026  
QY 2821 CCGTTACGAGTATCACTCACTTAAGAGACTCTTATGATTAAGAGGCTGATTTG 2880  
DB 3027 CCGTTACGAGTATCACTCACTTAAGAGACTCTTATGATTAAGAGGCTGATTTG 3086  
QY 2881 CTACGCTAGAGATTTGATTAATGATTTGATTTGATTTGATTTGATTTGATTTG 2940  
DB 3087 CTACGCTAGAGATTTGATTAATGATTTGATTTGATTTGATTTGATTTGATTTG 3146  
QY 2941 CTTTACCCCTAATGCTGATTAATTTGCAAGGCTTAAGGTTAATTTACTCTATCTCAAAA 3000  
DB 3147 CTTTACCCCTAATGCTGATTAATTTGCAAGGCTTAAGGTTAATTTACTCTATCTCAAAA 3206  
QY 3001 CTTATCCCTAATGCTGATTAATTTGCAAGGCTTAAGGTTAATTTACTCTATCTCAAAA 3060  
DB 3207 CTTATCCCTAATGCTGATTAATTTGCAAGGCTTAAGGTTAATTTACTCTATCTCAAAA 3266  
QY 3061 ACTGAGGCTCAGTACAGAGTACGCGGAAATATGATGAAGTAAATGACAAATCAAA 3120  
DB 3267 ACTGAGGCTCAGTACAGAGTACGCGGAAATATGATGAAGTAAATGACAAATCAAA 3326  
QY 3121 TGAAGTATGCAATGCTCAAAAGATCATTTGATTTGATTTGATTTGATTTGATTTG 3180  
DB 3327 TGAAGTATGCAATGCTCAAAAGATCATTTGATTTGATTTGATTTGATTTGATTTG 3386  
QY 3181 TACGGGTAACAGTTCTCAAGATATGTTAAATGATTTGATTTGATTTGATTTGATTTG 3240  
DB 3387 TACGGGTAACAGTTCTCAAGATATGTTAAATGATTTGATTTGATTTGATTTGATTTG 3446  
QY 3241 GAATCCGATGATTTTCTTAACCTCTCTCACTAATGATTAAGTAAATGATTTGATTTG 3300  
DB 3447 GAATCCGATGATTTTCTTAACCTCTCTCACTAATGATTAAGTAAATGATTTGATTTG 3506  
QY 3301 CGAAGTAAATGCGATGAGATGCTTAACCCAGTCTCAAGAGTAAACCTGCGGAGTAA 3360  
DB 3507 CGAAGTAAATGCGATGAGATGCTTAACCCAGTCTCAAGAGTAAACCTGCGGAGTAA 3566  
QY 3361 TCCGATGCTAAACCTCTCACTAATGATTTGATTTGATTTGATTTGATTTGATTTG 3420  
DB 3567 TCCGATGCTAAACCTCTCACTAATGATTTGATTTGATTTGATTTGATTTGATTTG 3626  
QY 3421 GCGTATTTCAAAACCAAGAGAGGCTCTTCACTCAAGAGGAGGAGCATTAAGAGGCT 3480  
DB 3627 GCGTATTTCAAAACCAAGAGAGGCTCTTCACTCAAGAGGAGGAGCATTAAGAGGCT 3686  
QY 3481 GGAATCTACTGCTCAGGTTGATTAATCAACCAAGAGGAGGAGGAGGAGGAGGAGG 3540  
DB 3687 GGAATCTACTGCTCAGGTTGATTAATCAACCAAGAGGAGGAGGAGGAGGAGGAGG 3746  
QY 3541 GGAATCTACTGCTCAGGTTGATTAATCAACCAAGAGGAGGAGGAGGAGGAGGAGG 3600  
DB 3747 GGAATCTACTGCTCAGGTTGATTAATCAACCAAGAGGAGGAGGAGGAGGAGGAGG 3806  
QY 3601 AGAAGATGAGCATCTCTGATTAACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3660

D	3897	AGAAATTGGCACTACTCTTGACTACCCAGAAACCTGCTGAAATGAACAGCAAAATCC	3868
Q	3661	CTCGTACTGGCTAATAAGAAATCGGCTAACTGAAAAAATAGATGAGGAGAAAATTTC	3720
D	3867	CTCGTTACTGGCTAATAAGAAATCGGCTAACTGAAAAAATAGATGAGGAGAAAATTTC	3926
Q	3721	ATTTCGGGGGGGTGATGATGATCTCTGCTGATCAACTCAATGTTGAGCACTCTATGT	3780
D	3927	ATTTCGGGGGGGTGATGATGATCTCTGCTGATCAACTCAATGTTGAGCACTCTATGT	3966
Q	3781	TGAAACAAGAGATGCTGTGTAAACGTCCAACTTCAAGACGAACTTCTCAAGAGGAGAGAT	3840
D	3987	TGAAACAAGAGATGCTGTGTAAACGTCCAACTTCAAGACGAACTTCTCAAGAGGAGAGAT	4046
Q	3841	ACTTAATGAACAACTGATATAGTTGATACCTGTTCTAAATTTATGGAATGAAAAAT	3900
D	4047	ACTTAATGAACAACTGATATAGTTGATACCTGTTCTAAATTTATGGAATGAAAAAT	4106
Q	3901	CTCTCCATTATTCATTAGATGATCATCACTGAACTTGCTCAAGCTCGGAGAAATGATAG	3960
D	4107	CTCTCCATTATTCATTAGATGATCATCACTGAACTTGCTCAAGCTCGGAGAAATGATAG	4166
Q	3961	CAATTCAACAGATTCGACATTCAATCGAAAAATGAAATATCACAGTGATTAATGAGA	4020
D	4167	CAATTCAACAGATTCGACATTCAATCGAAAAATGAAATATCACAGTGATTAATGAGA	4226
Q	4021	AAATACGAAATTAATCCAAAAACATGGAAGATCCCTTCTGTAAAGACATTGGATGA	4080
D	4227	AAATACGAAATTAATCCAAAAACATGGAAGATCCCTTCTGTAAAGACATTGGATGA	4286
Q	4081	ATTGAAACGTATAATGTTGATGAAATTTAGAAATGTGATTAATGACATGACATTGCTGA	4140
D	4287	ATTGAAACGTATAATGTTGATGAAATTTAGAAATGTGATTAATGACATGACATTGCTGA	4346
Q	4141	AGCTAATCCACTACACAGATGAAATTAATGATGTTCAAGTGAATAGAGTTTGTATATTA	4200
D	4347	AGCTAATCCACTACACAGATGAAATTAATGATGTTCAAGTGAATAGAGTTTGTATATTA	4406
Q	4201	TCATAGCATGTCACGAGCAAGAAAGAAATACACATTTGAGAAAGAGTTAACGAAAAAT	4260
D	4407	TCATAGCATGTCACGAGCAAGAAAGAAATACACATTTGAGAAAGAGTTAACGAAAAAT	4466
Q	4261	TGCTGTGATCTAAACATTTCATCTGATCAACTGATCCAGAGAACATCAGAGTTTAA	4320
D	4467	TGCTGTGATCTAAACATTTCATCTGATCAACTGATCCAGAGAACATCAGAGTTTAA	4526
Q	4321	TACTGTGTAACCAAGAGATCGAACCCCAAGAAAAAGAGGTGCTCATCACTGTGAATTT	4380
D	4527	TACTGTGTAACCAAGAGATCGAACCCCAAGAAAAAGAGGTGCTCATCACTGTGAATTT	4586
Q	4381	AAACAAAGATCGCAATCANGTCCATATGTTTCAAAGAAATGCTGAGACGTTTATAA	4440
D	4587	AAACAAAGATCGCAATCANGTCCATATGTTTCAAAGAAATGCTGAGACGTTTATAA	4646
Q	4441	CCCCAAGAGTATTTCATGCGCGTCGTCAACAAATCGACTATATGATCCGGGATGAT	4500
D	4647	CCCCAAGAGTATTTCATGCGCGTCGTCAACAAATCGACTATATGATCCGGGATGAT	4706
Q	4501	AAAGTCAATGATGCTGAACTAGAGAAATTTAGATCAAAAGATGTTTACGAAAGTTCC	4560
D	4707	AAAGTCAATGATGCTGAACTAGAGAAATTTAGATCAAAAGATGTTTACGAAAGTTCC	4766
Q	4561	AATTCGCCACGGTGTGAAGCTCATATCTCATGAGTTGGGTGATATCTGAGAAAATTTGATC	4620
D	4767	AATTCGCCACGGTGTGAAGCTCATATCTCATGAGTTGGGTGATATCTGAGAAAATTTGATC	4826
Q	4621	TCTCAAAAGTGTGTTGCGAAATCAAGTGTGTTTCCATGCAACAGACAAAAGAAAA	4680
D	4827	TCTCAAAAGTGTGTTGCGAAATCAAGTGTGTTTCCATGCAACAGACAAAAGAAAA	4886
Q	4681	ATTGATATATGACCTTTTATGTTATGTTCACTCTGTTATTAATCTTGTGACTATTAAGTT	4740
D	4887	ATTGATATATGACCTTTTATGTTATGTTCACTCTGTTATTAATCTTGTGACTATTAAGTT	4946

QY	4741	ATTGACATPATAAGTGTGTAATTAGGAATGACATTTCAACATTTAACAGTCGAGTGGC	4800
Db	4947	ATTGACATPATAAGTGTGTAATTAGGAATGACATTTCAACATTTAACAGTCGAGTGGC	5006
QY	4801	GTATCTAAATGCCCTCTATTACTCATATCCAAATTATGTCTTTCCTCTAAATCAGT	4860
Db	5007	GTATCTAAATGCCCTCTATTACTCATATCCAAATTATGTCTTTCCTCTAAATCAGT	5066
QY	4861	ACCTTTGAAGAAAACGATTCGTGTATTAAGAAACGTCGTCTGCTAATGGGTTAAACAGTC	4920
Db	5067	ACCTTTGAAGAAAACGATTCGTGTATTAAGAAACGTCGTCTGCTAATGGGTTAAACAGTC	5126
QY	4921	GGGTTTGGAAATGATCACAACACTATCAAAAGATATTTGAAAGACATTTGGTTTACTCAAGT	4980
Db	5127	GGGTTTGGAAATGATCACAACACTATCAAAAGATATTTGAAAGACATTTGGTTTACTCAAGT	5186
QY	4981	TTTACACAAAGATGAGTTTATTTTCCATGAAATATGAAAGAGGATCACTAATATTTAGG	5040
Db	5187	TTTACACAAAGATGAGTTTATTTTCCATGAAATATGAAAGAGGATCACTAATATTTAGG	5246
QY	5041	TTTATATGTTGATGATATTCCTATATGTTGGAAAGTCCAAAAAAGTTATGATAATTTGT	5100
Db	5247	TTTATATGTTGATGATATTCCTATATGTTGGAAAGTCCAAAAAAGTTATGATAATTTGT	5306
QY	5101	GGATCAATTTGAGATCATATTTTGAAGTTAAAGTCTTGAGGAAATATCAAAATTATCTGG	5160
Db	5307	GGATCAATTTGAGATCATATTTTGAAGTTAAAGTCTTGAGGAAATATCAAAATTATCTGG	5366
QY	5161	TATTTGAATTTCTGTAACCCGATCTGGTTATATTTTATCTCAAGAAAAATTTCTCAAGAA	5220
Db	5367	TATTTGAATTTCTGTAACCCGATCTGGTTATATTTTATCTCAAGAAAAATTTCTCAAGAA	5426
QY	5221	ATTACTTAAGGATTTCAAACCTAGATGCTCATATGAGGAAAAACATACCCTGGAATCCGAA	5280
Db	5427	ATTACTTAAGGATTTCAAACCTAGATGCTCATATGAGGAAAAACATACCCTGGAATCCGAA	5486
QY	5281	TGACAAATATGAAAGGTTGCAATAATTTGTGAAGACGTTATATCCAGAGATGATTTTGA	5340
Db	5487	TGACAAATATGAAAGGTTGCAATAATTTGTGAAGACGTTATATCCAGAGATGATTTTGA	5546
QY	5341	AAAGGTTCCGAAATGAGACATTTGCTTGACCTGATGCTAAAAAACTATACCAAGTGGTG	5400
Db	5547	AAAGGTTCCGAAATGAGACATTTGCTTGACCTGATGCTAAAAAACTATACCAAGTGGTG	5606
QY	5401	TGGCCTGCTTTATGCGCTGCCACAAACAACGTCACATATATCGGTGCTAGTGAATTC	5460
Db	5607	TGGCCTGCTTTATGCGCTGCCACAAACAACGTCACATATATCGGTGCTAGTGAATTC	5666
QY	5461	GTTGGGTTCTAAATCTGCCAATCCAAATGTCAGATTAATGAGAAATTAATTAATGTCT	5520
Db	5667	GTTGGGTTCTAAATCTGCCAATCCAAATGTCAGATTAATGAGAAATTAATTAATGTCT	5726
QY	5521	TAGGTATATCAAAAATATGACATGGATATCAACATTTAGTACAAAAGAAACAGATTGAATAT	5580
Db	5727	TAGGTATATCAAAAATATGACATGGATATCAACATTTAGTACAAAAGAAACAGATTGAATAT	5786
QY	5581	ACCAACCAAAATCATTTGTTATCGAATTTTCACAGATGCTCATTTSCACACAGATTTGA	5640
Db	5787	ACCAACCAAAATCATTTGTTATCGAATTTTCACAGATGCTCATTTSCACACAGATTTGA	5846
QY	5641	TAGAAAAATCTATTAATGAGAACTTTGATTTATGTCATGAGAAATTTGGTCAATGGCGAC	5700
Db	5847	TAGAAAAATCTATTAATGAGAACTTTGATTTATGTCATGAGAAATTTGGTCAATGGCGAC	5906
QY	5701	CAAAAAACAAACGCTCATAGACCAAAAGCTCAGACGCTGTGGAATGTGTGCTCTAAATTA	5760
Db	5907	CAAAAAACAAACGCTCATAGACCAAAAGCTCAGACGCTGTGGAATGTGTGCTCTAAATTA	5966
QY	5761	TACAAATGTTGAACCTATCGAATATAAAAAACATTTATATGATTTGGGTTTGAAGTAGG	5820
Db	5967	TACAAATGTTGAACCTATCGAATATAAAAAACATTTATATGATTTGGGTTTGAAGTAGG	6026



QY 5821 TAAGATACATGTCATCAAGACCAAGCTGTGATTAAGTTTGAGAAATATATTTG 5880  
 Db 6027 TAAGTATCATTTGTCTCAAGACCAAGCTGTGATTAAGTTTGAGAAATATATTTG 6086  
 QY 5881 TCACCCACATGACCAATAGATATCTGTATTAAGTTTGACCCCAATGATCAATGATA 5940  
 Db 6087 TCACCCACATGACCAATAGATATCTGTATTAAGTTTGACCCCAATGATCAATGATA 6146  
 QY 5941 AGTATTTCAATATCTGTATTAAGTTTGACCCCAATGATCAATGATA 6000  
 Db 6147 AGTATTTCAATATCTGTATTAAGTTTGACCCCAATGATCAATGATA 6206  
 QY 6001 TCTAAGTGTGTATTAATCAAGCAATGCTGTAGGGATGATTAAGCCGTTAGACTAGA 6060  
 Db 6207 TCTAAGTGTGTATTAATCAAGCAATGCTGTAGGGATGATTAAGCCGTTAGACTAGA 6266  
 QY 6061 AGATATCAACACCTGATTCATAAATGCAATACCGCAATAGATGATTTATCTTACT 6120  
 Db 6267 AGATATCAACACCTGATTCATAAATGCAATACCGCAATAGATGATTTATCTTACT 6326  
 QY 6121 ATTATCGTATGCTCAATCAGGAGGTGTGTTGTCGACTATTTGTCAGAAACTG 6180  
 Db 6327 ATTATCGTATGCTCAATCAGGAGGTGTGTTGTCGACTATTTGTCAGAAACTG 6386  
 QY 6181 ATCAATGAAATGATGTTATTAATGAGATGAGAAATTTTTCATCAATCAGGTGAT 6240  
 Db 6387 ATCAATGAAATGATGTTATTAATGAGATGAGAAATTTTTCATCAATCAGGTGAT 6446  
 QY 6241 GACAGAACCTAACTATTTGTTAGTATTAATTAAGGATGAGAAATGCAATACCGCAAT 6300  
 Db 6447 GACAGAACCTAACTATTTGTTAGTATTAATTAAGGATGAGAAATGCAATACCGCAAT 6506  
 QY 6301 TATCAACGAGATAGAGGAGAGGTTTCATATATATCTTGATTAATTAATCTTGTTCT 6360  
 Db 6507 TATCAACGAGATAGAGGAGAGGTTTCATATATATCTTGATTAATTAATCTTGTTCT 6566  
 QY 6361 AATTCACTATACCAACTAGAGCTGTACACGCTCAATCTCAGGTAAAGATTTATATT 6420  
 Db 6567 AATTCACTATACCAACTAGAGCTGTACACGCTCAATCTCAGGTAAAGATTTATATT 6626  
 QY 6421 CCATCA 6426  
 Db 6627 CCATCA 6632  
 RESULT 3  
 CAL251464/c 4872 bp DNA linear PLAN 08-JUL-2002  
 LOCUS CAL251464  
 DEFINITION Candida albicans eng1 gene for endo-1,3-beta-glucanase and pcal  
 retrotransposon.  
 ACCESSION AJ251464  
 VERSION AJ251464.1 GI:6562333  
 KEYWORDS endo-1,3-beta-glucanase; eng1 gene; retrotransposon.  
 SOURCE Candida albicans  
 ORGANISM Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 REFERENCE 1 Esteban, P.F., Garcia, R., Rios, I., Vazquez De Aldana, C.R. and del  
 Rey, F.  
 TITLE Cloning and characterization of the  
 endo-1,3-beta-glucanase-encoding gene CaENG1 in the yeast Candida  
 albicans  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4872)  
 AUTHORS del Rey, F.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-DEC-1999) Del Rey F., Departamento de Microbiología y  
 Genética, Universidad de Salamanca, Campus Miguel de Unamuno,  
 Salamanca, 37007, SPAIN  
 FEATURES  
 source 1..4872  
 /organism="Candida albicans"

/mol\_type="genomic DNA"  
 /strain="ATCC64385"  
 /db\_xref="taxon:5476"  
 711..4148  
 /gene="eng1"  
 711..4148  
 CDS  
 /gene="eng1"  
 /codon\_start=1  
 /evidence=experimental  
 /transl\_table=12  
 /product="endo-1,3-beta-glucanase"  
 /protein\_id="CA62579.1"  
 /db\_xref="GI:6562334"  
 /db\_xref="SPTREMBL:Q9UR03"  
 /translation="MLFKSVLSTLIAVALAENSAHCEIVTVTKTRHSINPLENYA  
 KLLPNNPTGILTTGPIVIVVAPPOODNVKQSESIQPOSSKAOQOQIOGSSVPS  
 IHDPTPTPTHTKIIVTSTTICNCKCTVKTSTIISHSSPTTTTTRKNEKSTADT  
 KSTFKALPSNGAVTLKTTISVKALITTSIVFSKLSSTSTSTSTSTSTSTSTSTST  
 PVSSQGVSTTTKTSINTSLITDAVITTEATLTPNSKSTFTNGSISFISTAVK  
 KSTVTSDIENNSONRSLSLITDQALSTLTITIKPTSLVPDEFSTNSRVSS  
 IKNSSSESSMYLEKSSKLSLSTSSLSASTSESLASTSDSLSHSSSS  
 VELSSLSSEADSSSSSESEVETGSSDETSTNSGDLFKADINAPVAREIPLTI  
 PAGVDNNGKPIGNNKRYTTLILGNDPWPYPLPYGLYWKSTSYGPAVQHNVSRLF  
 GSINTNNKVASIYRPTNNALIRSATSPSKDSMHMKYSQVLESLATYLTSSSND  
 SNLIDPLVQMGFVYGLYNGULPLNLSLFVYKDLSTSDALSVALKPAITLNL  
 VQMLIVTLPLDKDDEKLEVEDFYMLKSKSPVDGLIQLVALPEQNDNDKYDALAGM  
 YVYGAVSGSVSGTAAASYKFSYTAGKSSNNPIYFALPHMDSLTGSLADLANGLT  
 VSTTKEMTGFLNLEPSETINODEVLEPWTENVTGSLTYKQLELALANGL  
 ADIATVKNMNSVNSKSKVLDPKQIILVSEIIODEVTQDALNMDAKVFNQV  
 KQYVPLMVDTRKGVSTSTSAQDGPADPGSAVYNDHDHRYVFTIHAALIVGVYKX  
 GGTMAQSKDWNLSLVDRSNSADDTTPYVSRMPDPFSGHSMAGLPTTYKISSSS  
 ESLHFAALIKLVKVVQDSKEARGDLNLSIARSKNMFYKSDTVLPAKQILEKV  
 SGIFENKVDYTFTEGTPADHPEVYGHMLPTBSSLVRTSYVQSEWKQIDAFI  
 DNVDGWTGILRLNQALFDPKSSYEFASNNMDDXVLDNGOSRTWSLAPAGALNAS"  
 4441..4720  
 /transposon="pCal"  
 complement(4441..4720)  
 /note="left long terminal repeat of retrotransposon pcal"  
 ORIGIN  
 Query Match 6.7%; Score 432; DB 8; Length 4872;  
 Best Local Similarity 100.0%; Pred. No. 4,3e-59;  
 Matches 432; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5995 TAAGTGTCTAAGTGTCTAATTCAGAGCTTCTTGAGGGATGATTAACGGTTAA 6054  
 Db 4872 TAAGTGTCTAAGTGTCTAATTCAGAGCTTCTTGAGGGATGATTAACGGTTAA 6114  
 QY 6055 CCTAGAAGATTAATCAACACTGATTAACAATGCAATAGGCGAGATTAAGTATTC 6114  
 Db 4812 CCTAGAAGATTAATCAACACTGATTAACAATGCAATAGGCGAGATTAAGTATTC 4753  
 QY 6115 ATTACTATTAATGCTCAATGCTCAATGAGGAGGTGTGTTGTCACATAATTTGTCAG 6174  
 Db 4752 ATTACTATTAATGCTCAATGCTCAATGAGGAGGTGTGTTGTCACATAATTTGTCAG 4693  
 QY 6175 AATCTAGTAATGAATGATGCTTATTAATGAATGAATGATTTTTCATACACATATA 6234  
 Db 4692 AATCTAGTAATGAATGATGCTTATTAATGAATGAATGATTTTTCATACACATATA 4633  
 QY 6235 GGTGATGACAGAACTAACTAATTTGTAGTAAATTAAGGATGAATACCAATC 6294  
 Db 4632 GGTGATGACAGAACTAACTAATTTGTAGTAAATTAAGGATGAATACCAATC 4573  
 QY 6295 CGAATATTAACGAATAGAGGAGGAGTTTCATATATATCTTGTAATTAATTA 6354  
 Db 4572 CGAATATTAACGAATAGAGGAGGAGTTTCATATATATCTTGTAATTAATTA 4513  
 QY 6355 CGTTCTAATTCATACCAACTAGACGTGTACAGCTCAATCTCAGGTAAAGAAAGTT 6414  
 Db 4512 CGTTCTAATTCATACCAACTAGACGTGTACAGCTCAATCTCAGGTAAAGAAAGTT 4453  
 QY 6415 TATATTCATCA 6426

Db 4452 TATATTCATCA 4441

RESULT 4  
LOCUS AF030556  
DEFINITION Candida albicans retrotransposon pCal LTR, linear PLN 01-JUL-1998 cds.  
ACCESSION AF030556  
VERSION AF030556.1 GI:3273502  
KEYWORDS  
SOURCE Candida albicans  
ORGANISM Candida albicans  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; microsporitic Saccharomycetales; Candida.

REFERENCE  
AUTHORS Goodwin,T.J.D. and Poulter,R.T.M.  
TITLE Temperature- and strain-dependent expression of the pCal retrotransposon of Candida albicans

JOURNAL  
REFERENCE Unpublished  
AUTHORS Goodwin,T.J.D. and Poulter,R.T.M.  
TITLE Direct Submission  
JOURNAL Submitted (21-OCT-1997) Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand

FEATURES  
source  
1..408  
Location/Qualifiers  
/organism="Candida albicans"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:5476"  
/transposon="retrotransposon pCal"  
1..280  
281..312  
primer\_bind  
/standard\_name="minus-strand primer binding site"  
/note="Perfect match to internal region of Candida albicans transfer RNA-Ary-UCU"  
stem\_loop  
324..396  
398..5408  
/note="Start of GAG ORF"  
/codon\_start=1  
/transl\_table=12  
/product="GAG"  
/protein\_id="AAC24759.1"  
/db\_xref="GI:3273503"  
/translation="MSSA"

ORIGIN

Query Match 6.2%; Score 400; DB 8; Length 408;  
Best Local Similarity 98.8%; Pred. No. 1.1e-53;  
Matches 403; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGTGGTTCGCACTATTTGTGTCGAAACCTGATCAATGAAATGATGTTTATGA 60  
Db 1 TGTGGTTCGCACTATTTGTGTCGAAACCTGATCAATGAAATGATGTTTATGA 60  
QY 61 GATGGAATAATTTTCCATCAGATCAGTGTGACAGAACTAATATATTGTATGT 120  
Db 61 GATGGAATAATTTTCCATCAGATCAGTGTGACAGAACTAATATATTGTATGT 120  
QY 121 ATAAATTAAGGATTAATACCAACATCCGAAATATCAAGAGATGAAAGGAGGATT 180  
Db 121 ATAAATTAAGGATTAATACCAACATCCGAAATATCAAGAGATGAAAGGAGGATT 180  
QY 181 TCAATATATATCTGTGATATATTAATCTGTTCTAATTCATATACAACTAGAGCTGT 240  
Db 181 TCAATATATATCTGTGATATATTAATCTGTTCTAATTCATATACAACTAGAGCTGT 240  
QY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTTATATTCATCAGATTAGAAATCGATAGAT 300  
Db 241 ACACGCTCAATCTCAGGTAAAGAAAGTTTATATTCATCAGATTAGAAATCGATAGAT 300  
QY 301 AATCATTTGTCGCCAAATTAAGGTTGTATTAATTCAGTCCGAGATTGTATATGATT 360

Db 301 AACCATTTGCCCCCAATTAAGGTTGTAATTAATTCAGTCCGAGATTGTATATGATT 360

QY 361 GATAGTTTCGAAGTTTGAAGGTACAGATTTCAACAGATGATCCCG 408  
Db 361 GATAGTTTCGAAGTTTGAAGGTACAGATTTCAACAGATGATCCCG 408

RESULT 5  
LOCUS DHA439551  
DEFINITION Debaryomyces hanseoni var. hanseoni LTR-retrotransposon Tdh2.  
ACCESSION AJ439551  
VERSION AJ439551.1 GI:20152518  
KEYWORDS gag gene; integrase; long terminal repeat; LTR; pol gene; protease; reverse transcriptase; RNaseH.  
SOURCE Debaryomyces hanseoni var. hanseoni  
ORGANISM Debaryomyces hanseoni var. hanseoni  
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Debaryomycetes.

REFERENCE  
AUTHORS Neuveglise,C., Feldmann,H., Bon,E., Galliardin,C. and Casaregola,S.  
TITLE Genomic evolution of the LTR-retrotransposons in hemiascomycetous yeasts

JOURNAL  
REFERENCE Unpublished  
AUTHORS Neuveglise,C.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-2002) Neuveglise C., Genetique Molculaire et Cellulaire, Inra, Thiverval-Grignon, 78850, FRANCE

FEATURES  
source  
1..5928  
Location/Qualifiers  
/organism="Debaryomyces hanseoni var. hanseoni"  
/mol\_type="genomic DNA"  
/strain="CBS767"  
/variety="hanseoni"  
/db\_xref="taxon:58641"  
1..5928  
repeat\_region  
/transposon="LTR-retrotransposon Tdh2"  
1..379  
380..394  
primer\_bind  
411..1421  
411..1421  
411..1421  
gene  
/gene="gag"  
/gene="gag"  
/codon\_start=1  
/transl\_table=12  
/protein\_id="CAD29536.1"  
/db\_xref="GI:20152518"  
/db\_xref="SPTREMBL:Q8TFK1"  
/translation="MSEQPTVDPTLVGHAKTVNNRVYESSNEEMGSKTYKIHHSKS  
SDVINKPSSSLKFNQKEIVDNAIEQAKELVOCYFEKIPFVHVDVSTLLDLMAN  
KLSNKGKGLLINDPTIDYGVKKS SVYCNNDLQIFPVRELLSGLFHTS  
ALOKYRPEVFSINVTYKYIRDNEVFVSEILKLINFYIHAMDNPMDVTDIATAT  
VGTGEILDELINDPKLDLYRILLYRKLTKNKEAFIDELTPAVLEFPNNLL  
NDQILVNLKRYTRFKRYFOEKTDYNSNFDDNKKKQSKNNKYEHRRNNSKXND  
EN"  
1422..5444  
/gene="pol"  
<1422..5444  
/gene="pol"  
/note="contains protease, integrase, reverse transcriptase  
and RNaseH"  
/codon\_start=1  
/transl\_table=12  
/protein\_id="CAD29537.1"  
/db\_xref="GI:20152520"  
/db\_xref="GOA:O8TFK0"  
/db\_xref="SPTREMBL:Q8TFK0"  
/translation="QIGSHNTDIKUSFSINNISSDPFINNOYTIYNNKTKDSEKNY  
FDSGASVTVVDIKQLINIKEDLSESLFTVSNVPIYAAKEDLVYIKNKCFITN  
NAVYIPGIQSNLSIFNDLNGPLFLPNDYLPFLKSGNKGCFIARQDTSNLFNNI  
RSRDFSMKLSRKIIDELNNMOKIINKKTLNYPNNKYIKISNNHETIANPFRKRD.

MYHLMGNHMSLESKYLKSGHIMSESEITASBEERYSCNECIAINSKOSHHHT  
 FTAPEPLRHSBTDLGIISHRGRKYITITLIDEXGYANTTMSHKSLODLPEKIRI  
 MNKRVADVNAEFTDNLMEFTKQDLAIEKDELSISPEALFISERNRSIOP  
 IRKALLPQDTRTLLPRIVDYVYINMTPVRSGSCCPALFISGTHPFIIVP  
 GLDVIVKSLSTEAKRYGVNSKSPCLFGFIYGVDVHYKVLISTDPFIIVP  
 NITLKMNDLKYVLLKSLDYLODKADIDISGLDTRDEKSLIAEHNOPTIE  
 CNEYDQESLVNSSTIODVAOHFERCHESNSLQMSPVSNMTWTDNDYDIT  
 EETMEVNSPBDTNETTDSNITKTNDNSNSOCSNCHOTLPHSADKND  
 EINKNDGNTLSLESSTLSKPTTQNNVRSKCAIVEYVERPKITSSKDR  
 VKDPHTKLSNRSKSKINDLQSHARKVNSLFTIYVITMSKRLNSOSKSN  
 SLVTRKINIVYKVDLTDNMKQSIQRELDTPKYEYTVVKNVNPPIPTVHTH  
 KINLKEVQYKRCVVOGFRQJANSHYDTSKVSFVLSILRLTALVAYEPIH  
 LDISAVYADIDYEKSIKVEPPSGNIDSGKQMLNSKVGMOAGVWOCCTKYL  
 MDLNEPDIAGMFCYKGEENKULIVLYVDMFLTSNITINDFLBELAKHEDK  
 YPADISEFLIEFIOAGVYLSOHNFINSVYKKNLNNNGKYIPIIKENKSEFO  
 LKSEFTNBEVYKESLNEBEDKELQSVGSLMBAANTRPDISFANQUSNNQNT  
 SVDLKLIYCLRYKQVOTISFSLERYRNFPSHKSFPILQTSDESPFATKDRSITGY  
 TVYLNGLINMSTIRQKVIDTSASACEINALSAVRSILKSKROAILDLINVIDEITLF  
 EDNAVIANONNEGTSYGRWVDIKLFIQVSEGLIKLKYVNTSINIADMLTKALS  
 RKLFEHLRSILFERDLNKE"

LTR  
 ORIGIN

Query Match 5.7%; Score 363.8; DB 8; Length 5928;  
 Best Local Similarity 52.5%; Pred. No. 2.8e-48;  
 Matches 872; Conservative 0; Mismatches 777; Indels 12; Gaps 3;

4444 CAAGAGTATTACATGCGCGTGTCAACAAATGACATATATAGTCCGGAGATGATAA 4503  
 3824 CACTGGAAAAATTAATACGTATATCGAAAGTCAATTAACAGATTAACAATGGAAACA 3883  
 4504 GTCAATGAATGCTGAAGTGAAGAAATTTAGATCAAAAGATTTTACGAAGAAGTTCAT 4563  
 3884 GTCAATACACGCGAATAGATATTTAAGAAATATGAAGATATACGGTTGGAAAAA 3943  
 4564 TCCCAACCGGTGGAAGCCATATCTATGGGTTGGTACATCTGAGAAATTTGATTTCT 4623  
 3944 TCTTAAAAATGCAACCTATTCACATCTGGGTTCTATACATCACTAAATTAACGACT 4003  
 4624 CAAGAGTGTGTCGGAATCACTGTGTGTCATGCGCAACAGCAAAAGAAATTAAT 4683  
 4004 CAAAGAGTTCAGTAAATCACTGTGTGTCATGCGCAACAGCAAAATTAATGA 4063  
 4684 GGATTAATGACCTTTTATGTTAGTTACCTGTTATAGATCTTGACTATTAAGATTAT 4743  
 4064 ACACATATATACCTTGAAGGTGTCATCTGCTGTGATTAATTAATCCAAATTTCTTACT 4123  
 4744 GACATTAATAGTGTGAATTAAGATGACATTAATCAATTTAGACGTCGAGTGGGGTA 4803  
 4124 TACAGCGATAGAGTTGAATTAAGATGAGCGATACATCTTGATATATCTTCGCAATA 4183  
 4804 TCTAATGCTCTTATTAATCAATCAATTTATGCTCTTCTCTTAATCAGTAACC 4863  
 4184 TTTACATGACATGACATGACATGACATTAATTTGTAACACACCACTGGATGAA 4243  
 4864 TTTGAGAAAAACCTGTGTGTTATGAAAGTCTGCTATGGGTTAAACAGTGGG 4923  
 4244 TATTATCTTGTGTAATGTGGCAATTAACAAATCTTTTATGGAAGAAACAAGAGG 4303  
 4924 TTTGAGATGATCACTATCAAAAGATTTGGAAGACATGGTTTACTCAAGTTT 4983  
 4304 GTATATGTGTATCAATGATTAAGTCTTATGATCTTATTTGCAACCTGATAC 4363  
 4984 ACAATATGATGTTTATTTACATGATTAAGAGGAGATCAGTAATATATTTAGGTT 5043  
 4364 TGCATTTGCGGAAGTTTGTAAATATTTTGTGAAAAATAGAAGCTCAGTGTGCACT 4423  
 5044 ATATGTATGATATTTCTTATGTTGAGATTCACAAAGATTAATGTAATTTTGTGA 5103  
 4424 ATATGTATGATATTTTATGTTTACTGCTTAATTAATTTCTTAACGATTTCAACT 4483  
 5104 TCAATTTGAGATCTTTTGAAGTTAAAGTGTGGGAAATATCAATTAATCTTGGTAT 5163

4484 TGAATCTGCTAAACATTTGCACTTAAATATTTGCTGACATATCTGAATCTTTAGAAAT 4543  
 5164 TGAATTTGCTAAACCGAATCTGTTATATTTTATCTCAAGAAAAATTTCTCAAGAAAT 5223  
 4544 TGAATTTGCTAAATTTGCAAGGGGATATGATTAATCAACAACAAATTTCTGAAATCTGT 4603  
 5224 ACTTAAGATTTGAAACTGATACATATATGGAAGAAAAATACCTGATTCGCAATGA 5283  
 4604 AATTTAAAAATTAATTAATTAATTTTATGGAAGATATTTCAATTAATTTTAAAG 4660  
 5284 CAATATGAAAGATTTGCAATTAATTTGCTGAAACGTTAAATCCAGAGATGATTTGAAA 5343  
 4661 GAAAAATTAATTTCCGAAA-----CTCACTAAAGAGTAATTTACTAATGACGAAGT 4714  
 5344 GGTTCGAATGACATTTGTTACCTGATGTGTAATAAATTTTACCAGAGTGTGTGG 5403  
 4715 CATTAATAAATGAATCATTTGTTAAATGAAGAAATTAAGCTATATCAATCAAGTGTGG 4774  
 5404 CTGCTTTTATGGGCTGCCAACAACAGCTCCAGATATATCGGTGATGATTCGTT 5463  
 4775 TTGCTATTTATGGGAGCTAATTAACAAGGCGCAGACATAGTTTGTGTAATCAAT 4834  
 5464 GGGTTTAAATCTGCAATTCAAATGCTCAATGATTAAGAAATTTGATTTTGTCTTNG 5523  
 4835 GAGCTCAATTAATCAAAATCCGACATCAGTGCATCTTGAATAATGATATATGTTTACG 4894  
 5524 GTATATCAAAATATAGATGGGATATCAATTTGATGATCAAAAGAAACAGATTAATTAAC 5583  
 4895 TTATGTAATAACAACATATCTTTTATGATTAAGATTAAGCGAATTCAGATTTTCAATTA 4954  
 4955 AAGGAGATCATTTATCATTTCAATCTTTTCAATGATCTTTTGCAACCAACAAAGATAG 5014  
 5644 AATATCATTTAGGCACTTTGATTTATGTAATGAATGAATTTGTGCAATTTGGCGCA 5703  
 5015 AAGATGATTAACGTTATATACGTTATTAATGATTAATGATTAATTTGCTTACCAT 5074  
 5704 AAAAAAAGCGTATAGCAACAAAGCTCAGCAGCTGTGAAATGTTGCTCTAATTTATAC 5763  
 5075 AAGGAGAGGTTATCACTGATATGTTCTGACGCTTGAAATCAACAGCTACATTTACAG 5134  
 5764 AATGTTGAAAGCTATGAAATTAATAAATTTATGATTTGATTTGGGTTTGAAGTACGTA 5823  
 5135 TGTGAGAGTACCTTAATCAAGACAGCTATTTAGATCTAAATTTGATTAATGATGA 5194  
 5824 GATACATTTGATCAAGACAAACCAAGCTGTGATTAAGTTTGAAGAAATTA---CTATTG 5880  
 5195 AATTAATTTATTTGAAGACAAACCGCGAGTTATAGGAATCTGTAATTAAGAGTACTTC 5254  
 5881 TCACCAATTCGACCAATGATATCTGATTAAGTTTCTACGCCAATTTGATCAATGATTA 5940  
 5255 ATATTCAGGCGATGCTGATATCAATTAATTAATTTATCAAGAGTTGTTCTGAAG 5314  
 5941 AGATTTTCAATCTGATGAGACAAATGATAATTACGCGGATTTGATGATGATGAG 6000  
 5315 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5374  
 6001 TCTAATGCTGTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6060  
 5375 ATTAAGCGGAATTAATTTGAAATATTTGAAATCTTGTGCTACATATTTGAAAGAAATTA 5434  
 6061 AGATTAATCAACCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6101  
 5435 TAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5475

RESULT 6  
 AF078809  
 LOCUS  
 DEFINITION  
 Candida albicans Tca4 retrotransposon reverse transcriptase (pol)  
 gene, partial cds.  
 ACCESSION  
 AF078809

```

VERSION      AF078809.1  GI:3377678
KEYWORDS
SOURCE       Candida albicans
ORANISM      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
              Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE    1 (bases 1 to 1470)
              Goodwin, T.J.D.
              Tc4, a Ty1-copia retrotransposon from Candida albicans
              Unpublished
              2 (bases 1 to 1470)
              Goodwin, T.J.D.
              Direct Submission
              Submitted (16-JUL-1998) Department of Biochemistry, University of
              Otago, Cumberland Street, Dunedin, New Zealand
FEATURES
  source
    1..1470
    /organism="Candida albicans"
    /mol_type="genomic DNA"
    /strain="SC5314"
    /db_xref="taxon:5476"
    <1..>1470
    /gene="pol"
    <1..>1470
    /gene="pol"
    /codon_start=2
    /transl_table=12
    /product="reverse transcriptase"
    /protein_id="AAC28090.1"
    /db_xref="GI:3377679"
    /translation="GRSVHPKRYLNAIVKIDYNOKWQSEEEIEKRNQVYTV
    EKPRNVPLKTMVHTYKNDLKNHYKSRVYGVNMYENEDPDHAFSSVPLT
    SIKLSAIVANNIMHOLDIASVYLAASLEDGVIPEPRGGEVPGASMLHSV
    YGLQSAHMYSHKRVYLEANGKQTLHNGIIPKYNENDVIVSYVDPMIAIN
    EKIIKEFVAMLETYFQIQREBATEYIGIQRKPPDYITDQIPLEKVIATENIDS
    YGKQIPILPKDINVYKQLRSSQINDVKEKPKLINSYKTIYDIEHEHPRK
    ESPQAEPLSAKGIKLYQSAVGSILMASMNRPDIAFVNOLGAKCAPVDWKLRY
    CLRYKMDKLEKRGKRLNNKSDFIICFSPSAPAPDLDRKSIYGISIFVNGMLV
    AMATKOKIITHSSAACEMLANITYILKAPDLRTIEDL"
  repeat_region
    <1..>1470
    /note="Tc4 retrotransposon"
    /rpt_family="Ty-1-copia"
    /rpt_type="dispersed"

ORIGIN
Query Match      5.3%; Score 339.2; DB 8; Length 1470;
Beef Local Similarity 54.4%; Pred. No. 3.4e-44;
Matches 790; Conservative 0; Mismatches 578; Indels 84; Gaps 2;

4423 TGGTAGAAGCGTTATTAACCCCAAGAGGTATTACATGCGTCTGCAACAAATCGACTA 4482
      |||||
      1 TGGAGATCAGTGCATCCACCTAAAGATATTAAATGCCATGTGTAAGAAATAGATTA 60
      |||||
      4483 TAAATCCCGGATGATGAATGCAATGAATGCACTAGAGAAATTAGATCAAGA 4542
      |||||
      61 CAATCAAAAAGATGGCTCAAGTATGGAAGAAAGAAATCGAAAATTAGGCTAACCA 120
      |||||
      4543 TGTTCAGAGAAAGTTCATTCACCGCGGTGAAGCTTATCTATGCGTGGGTACA 4602
      |||||
      121 AATTTCACCGTTGAAAAACCAAGAACGTTGTCATGTAACCAATGAGGTACA 180
      |||||
      4603 TACTGAGAAATGATCTCTCAAGGTGTTGCGAAATCAGCTGTGTTCCTCANG 4662
      |||||
      181 TACTTACAAACCAATGACCTCAAAATCATATTAACAAAGCGTGGGTGATAGGG 240
      |||||
      4663 CAACAGACAAAGAAATGATGATGACCTTTTATAGTGTAGTTCACCTGTTATAGA 4722
      |||||
      241 AAACTATATGTCGAAAATCGTATTTGATCCCATGCCATTCCTCCCGGTAGTACA 300
      |||||
      4723 TTTTGACTATTAAGTTTATGCAATATAGGTGTGAATTAGAAATGAATTTACA 4782
      |||||
      301 TCTCAAGATATACGACTATTATCTGCATAGCTGTTAAAAATACTTGTTATGACCA 360

```

```

QY 4783 TTTAGCGTGAAGTGGCGGTATCTAAATGCTTATCTATCAATTCATCAATTTATGT 4842
DB 361 ATTGACATTCGCTCAGCTTATTTGAAGCCAGTTTGAGATGGAAGATATCTTGT 420
QY 4843 CTTCCTCTTAATCAGTACCTTTGAAAGAAACCAATTTGTTTATGAAACCTTCTGT 4902
DB 421 GAGACCAACCGGTGGTTTGAAGTTAAACCTGCTATAGTTGGCTTTTACCAAGTCTGT 480
QY 4903 CTATGGTTAAACAGTGGGTTTGAATGGTATCACACATCAACAAAGAGATTGGAGA 4962
DB 481 GTACGGTCTTAGGACAGAGTGGCCATTAATGGTATCTACATTTTGAAGATGTTGGAGCC 540
QY 4963 CATGGTTTATCTCAAGTTTACACAAATGATGTTTATTCACATTAATATGAGAGGG 5022
DB 541 AAATGTTTAAACAAACATACACAAATGATGATTTTGAAGAAATATGAAATGG 600
QY 5023 ATCAGTAAATATTTAGTTTATATGTTGATGATATCTTATGGTGGAGTTACAAA 5082
DB 601 AGATGATTAATATGAGTGTATATGAGAGATGTTTATGATAGCTAATATGAAAA 660
QY 5083 AGTATGATATATTTGATGATCAATGAGATGATTCATTTGAAGTTAAGTGTGGA 5142
DB 661 GATTATTTAAAGTTTGTGTATGCTATGCTGAAACATATTTCACTTACATATTTGGCGA 720
QY 5143 AATATCAATTAATCTTGTGTTATGAAATTTGTAACCAATCTGTTATATTTATCTCA 5202
DB 721 GGTACGAGATCTAGGAATATACGTTTGAAGAACACCTGATGATACAGTTTACCA 780
QY 5203 AAAAAATTTCTCAAGAAATTAATTAAGATTTTCAACTAGATACATATGAGAAAA 5262
DB 781 AATACCATTTCTGAAAAATTTGGTGCATCACTTCAATTTCAATATGATTAAGAAAA 840
QY 5263 CATACCCTGATTCGAATGACAAATATG----- 5291
DB 841 CATTCCTATATCCCAAGAAATATTAATGTGTGTCAGCAATTTGAGAAATGCAAGTCAGAT 900
QY 5292 AAAAGTTGCATATTTTGTGTAAC----- 5317
DB 901 AATGATTTTGTAAAGTGAAGAACTTCAAAAGTTATCAATTCGTTGCTCAAAACCA 960
QY 5318 -----GTTAATCCAGAGATGATTTTGAAGTTCCGAATGAGAC 5358
DB 961 GTACATGATGATTAAGAAAGAAATGAGAAACATGATTTTAAAGATCACCAAGCTGA 1020
QY 5359 ATTGCTTACCTGATGCTGTAAGAAACCTATACCAAGTGTGGCTGCTTTATGGGC 5418
DB 1021 ACCCTTAAGCGGAAAGAAATTAATTAATACCAATCCGAGTGGAGCTGTTATGGGC 1080
QY 5419 TGCACAAACACAGTCCAGATATATCGGTGTAGTGAATTCGTTGGTTCTAATCTGC 5478
DB 1081 TTGATGAACATTAAGCCAGACTTATAGCTTTAGTGAATCAGCTCGAGCAAGGTGTC 1140
QY 5479 AAATCCAAATGTCATGATTAAGAAATGATTTATTTGCTTATAGCTATATCAAAAAATAG 5538
DB 1141 TCACCGAGATGTCGAGCTGAGAAAGATGATGATGATGCTTACGATTAATTAAGAAAGA 1200
QY 5539 CATGGATATCACTTGTGATCAAAAGAAACAGATTGAATATACCAACCAAAATCATTTGT 5598
DB 1201 TATGGACTTCAACTTGAATACAGAGAGAAAGATCAACAAAGTCAAAAGATTTTAT 1260
QY 5599 TATCGAATGTTCAAGTATGCTCATTTGACACAGATTTGATAGAAATCTATTAAGTGG 5658
DB 1261 TATTTGATGTTTCTGATGCTCATTTGCCCCAATTTGGACAGAAAGTCAATACCGG 1320
QY 5659 AACTTTGATTAATGTAAGAAATTTGGTGAATGGGGGACCAAAAAACAAACGGTCAAT 5718
DB 1321 GACAGACATTTTGTAAATGGAATCTATGTTGATGGGCAACAAAGAAACAAAAATAT 1380
QY 5719 AGCAGAAAGCTCAGAGCTTGTGAATGTTGCTTAAATTAACAAATGTTGAAGCTAT 5778
DB 1381 CACACATATTCAGCGCTGTGGAAATGCTTCATTAATTAATTAACATATTAAGAGCAAT 1440
QY 5779 CGAATTAATAAAA 5790

```

OY		4201	TCAATGACATGTCCAGCACAAGAATAACACTTCTGGAAAGAACTTAACGAAAAAAT	4260
Db		43596	TAAATATCATATATATATATTTTTATATATATACATATAATCGAACAAAAACGAATAAAGA	4385
OY		4261	TGCTGTGTACTAAACATTCACCTTGATATACACTGATCCAGAGAACAA	4307
Db		43656	AGAAGTATCATTTTATATATCTGTAAACCACCCTAATGTATAGAACTA	43702
RESULT 9				
AC140550				
LOCUS				
DEFINITION	Medicago truncatula clone mh2-54a24, WORKING DRAFT SEQUENCE, 2			
	unordered pieces.			
ACCESSION	AC140550	128101 bp	DNA	linear
VERSION	AC140550.19	GI:39752730		
KEYWORDS	HTG; HTGS PHASE1; HTGS DRAFT.			
SOURCE	Medicago truncatula (barrel medic)			
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots; rosids; eurosida I; Fabales; Fabaceae; Papilionoideae; Trifolieae;			

REFERENCE	1 (bases 1 to 128101)
Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,	
COOK, D., Kim, D. and Roe, B.A.	
Medicago truncatula BAC Clone mth2-5424	
TITLE	Unpublished
JOURNAL	2 (bases 1 to 128101)
REFERENCE	Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
COOK, D., Kim, D. and Roe, B.A.	
TITLE	Direct Submission
JOURNAL	Submitted (25-FEB-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
OK 73019, USA	
REFERENCE	3 (bases 1 to 128101)
AUTHORS	Shauli, S., Lin, S., Dixon, R., May, G., Summer, L., Gonzales, B.,
COOK, D., Kim, D. and Roe, B.A.	
TITLE	Direct Submission
JOURNAL	Submitted (12-DEC-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,	
OK 73019, USA	
COMMENT	On Dec 12, 2003 this sequence version replaced gi:38678580.
----- Genome Center	
Center: Department Of Chemistry And Biochemistry	
The University Of Oklahoma	
Center code:UOKNOR	
-----	
* NOTE: This is a 'working draft' sequence. It currently	
* consists of 2 contigs. The true order of the pieces	
* is not known and their order in this sequence record is	
* arbitrary. Gaps between the contigs are represented as	
* runs of N, but the exact sizes of the gaps are unknown.	
* This record will be updated with the finished sequence	
* as soon as it is available and the accession number will	
* be preserved.	
* 1 14715: contig of 14715 bp in length	
* 14716 14815: gap of unknown length	
* 14816 128101: contig of 113266 bp in length.	
FEATURES	Location/Qualifiers
source	1..128101
	/organism="Medicago truncatula"
	/mol_type="genomic DNA"
	/db_xref="taxon:3880"
	/clone="mth2-5424"
	/clone_1db="Medicago truncatula BAC library H2"
ORIGIN	
Query Match	1.6%; Score 105.6; DB 2; Length 128101;
Best Local Similarity	47.3%; Pred. No. 1.2e-07;
Matches 352; Conservative	0; Mismatches 389; Indels 3; Gaps 1;
QY	4470 ACAAAATCGACTTAATGATCCGGAGATGAATAGCTGTAAGTGAAGAAAT 4529
Db	69336 ACAAGAAGCTTTAAGACTTAATTTGGAAAAGTGATAGTTGAAGAACTACATCA 69455
QY	4530 TTAGATCAAAAGATGTTTACGAGAGAGTCCAAATTCACCGGTGGAAGCTATATCTCA 4589
Db	69456 TTGAGAGAAATCAACCTGAGAGCTAGTATTACTTGCCAGCAAGAAAAAGAAAGTTGATG 69515
QY	4590 TGGGTGGGTGATACATGAGAAAATTGATCTCTCAAAAGAGTGTGTTCCGAATCACTT 4649
Db	69516 TCAATGGGTTTACGAGTGAATTAAGAACTGATGATAGTAAAGTTCTTAGACCAAGCA 69575
QY	4650 GTGTGTTCATGCGACAGACAAAAGAAAATTGATTTGACCTTTTAGTGTGATTT 4709
Db	69576 GATTGGTGGCTAGAGGTTTCTTCAGAAAGCA--GACTTAGACTTAATGAAGTGTG 69632
QY	4710 CACCGTTATGATCTTGTGACTTAATGATTATTTGCAATATAGGTTGAAATTAGAA 4769
Db	69633 CTCGAGTTGCTAGAAATGAACCTGAGAGTTAGTAAGTGTGATTTGCTGTGAAGAT 69692
QY	4770 TGAACATTCACATTTAGAGCTGAGAGTGGGATCTTAATGCTCTATTACTCATTC 4829
Db	69693 GGTAACCTACCAATTTAGATGAATGAAGTACAGATTTCTTAATGGTCCACTGAAGAGTGA 69752

QY	4830	ATCCAAATTATGCTTTCCTCCCTAAATTCAGTACCTTTGGAAGAAAAACCTGGTGTAT	4889
Db	69753	TGTTTGTGCACACACACACACAGATTTGGAATTTGCTGGGAAGAAAATATGCTGTAAAC	69812
QY	4890	TGAAAGCTTCTGCTATGCGTTTAAACAGCTGGGTTTGGAAATGATATCACACTATCAAAA	4949
Db	69813	TGCACAAAAACACTATATGATATTAATAATAGCTCCCTGAGACATGGAACAGAAAATTTGTC	69872
QY	4950	GACATTTGGAGACATTTGGTTTACTCAAGTTTACACANTATGGTTATTTTCATTTG	5009
Db	69873	AAAGTTTTCATCCAAATAGCTTTTAAATATTTGTTAGTTGATGTTTGGATATATGTTTACA	69932
QY	5010	AATATGGAAGGATCAGTAATATATTTAGGTTTATATGTATGATATATCTTATAGTTG	5069
Db	69933	ACCTTAGTGTATGAGGACACTGTTATATATTTGCTGATGTTTATATGCTTACTCATCAGT	69992
QY	5070	GAGGTCACAAAAAGTATTTGATTAATTTTGGATCAATTTGAGATCATTTTGAAGTTA	5129
Db	69993	GTTAGCTTACATCATAGATTTGAGAACGTTTAAAGAGAAATGAAAGTTGAGTTTGAATGA	70055
QY	5130	AAAGTTTGGTGAATATATCAATTAATTTCTTGTTATTTGTAATTTGGTAAACCGAATCTGTT	5189
Db	70053	CTGATCTTGGCGAGCTTTCATCTTCTTGGAAATGGAATTTTGTGAATTTGAAAGCTGGGA	70112
QY	5190	ATATTTTATCTCAGAAAAATTTTC	5213
Db	70113	TAGTATGATGATTCAGAAAAAATATTC	70136
RESULT 10			
LOCUS	AL935272	146275 bp	DNA linear HTG 30-OCT-2003
DEFINITION	Danio rerio clone CH211-266k22, WORKING DRAFT SQUENCE, 2 unordered pieces.		
ACCESSION	AL935272		
VERSION	AL935272.16	GI:37936458	
KEYWORDS	HTG; HTGS PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Danio rerio (zebrafish)		
ORGANISM	Danio rerio		
	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
	1 (bases 1 to 146275)		
	Andrew R.		
	Direct Submission		
	Submitted (29-OCT-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk		
	On Oct 23, 2003 this sequence version replaced gi:35762972.		
COMMENT	----- Genome Center		
	Center: Wellcome Trust Sanger Institute		
	Center code: SC		
	Web site: http://www.sanger.ac.uk		
	Contact: zfish-help@sanger.ac.uk		
	----- Project Information		
	Center project name: ZC26KX2		
	----- Summary Statistics		
	Assembly program: XGAP4; version 4.5		
	Chemistry: Dye-terminator; 100% of reads		
	Consensus quality: 145872 bases at least Q40		
	Consensus quality: 145995 bases at least Q30		
	Consensus quality: 146056 bases at least Q20		
	Insert size: 146175; sum-of-contigs		
	Insert size: 155501; 1.5% error; agarose-fp		
	Quality coverage: 9.35x in Q20 bases; sum-of-contigs quality coverage: 9.13x in Q20 bases; agarose-fp		
	-----		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 2 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		





Db 56113 GAAAGAGCTCTGTATGGCTTAAACAGACCTAGACATGGTATGAAAGATTACATA 56054  
 QY 4951 AGATATGGAAGACATATGTTTACTCAAGTTTACACATGATGTTTATTTACATTTGA 5010  
 Db 56053 ATTCTGACCAACATGTTATATGACAAAGGCGATGACAAACCTGTTTGTAAAGAA 55994  
 QY 5011 ATATGAAGAGGATCAGTATATATTTAGCTTATATGATGATATTCCTATGTTGG 5070  
 Db 55993 AAAAGGAGTAACTTATATGATGCTCAA---ATTATGTTGATGACATTTGTTGGTGG 55937  
 QY 5071 AAGTTCAAAAAGTTATTTGATATTTTGTGATCAATGAGATCATTTTGAAGTTAA 5130  
 Db 55936 CATGTGCAACCAAGATGATGGAACAATTTGTTGAACAATATCTGATTTGAATGAG 55877  
 QY 5131 AGCTTTGGTGAATATCAATATTTCTGTTATTTGAAATTTGCTGTTGTTGTTGTTA 5190  
 Db 55876 TCTGTATGTTGATTAATTAATTTCTGTTGAGCTTCAAGTAAACAGATGAAATCTCT 55817  
 QY 5191 TATTTTATCTCAAGAAAATTTCTCAAGAAATTTACTAAGATTTTCAACTAGATACCT 5250  
 Db 55816 ATTATGACACAAAGTATAGTATGTAAGGATTTGTTAAAGTTGCTTTGAGATGC 55757  
 QY 5251 ATATGGGAAAAACATACCTGCGATTCGGAATGACAAATATGAAAAGTTGCAATTAATCG 5310  
 Db 55756 AGTTCACAAAAGACACCTGCTGCACTCATATCACTGCAAAAGATGAGAAAGGTAC 55697  
 QY 5311 TGA 5313  
 Db 55696 TGA 55694

RESULT 12  
 AP006094/c 127731 bp DNA linear PLN 22-JUL-2003  
 LOCUS Lotus corniculatus var. japonicus genomic DNA, chromosome 4,  
 clone:UJ79H01, TM0172, complete sequence.  
 ACCESSION AP006094  
 VERSION AP006094.1 GI:29122733  
 KEYWORDS HTG.  
 SOURCE  
 ORGANISM Lotus corniculatus var. japonicus (Lotus japonicus)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Loteeae;  
 Locust.  
 REFERENCE 1  
 AUTHORS Kaneo, T., Asamizu, E., Kato, T., Sato, S., Nakamura, Y. and Tabata, S.  
 TITLE Structural analysis of a Lotus japonicus genome. III. Sequence  
 features and mapping of sixty-two TAC clones which cover the 6.7 Mb  
 regions of the genome  
 JOURNAL DNA Res. 10 (1), 27-33 (2003)  
 MEDLINE 12579290  
 PUBMED 12593552  
 REFERENCE 2 (bases 1 to 127731)  
 AUTHORS Sato, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-DEC-2002) Shusei Sato, Kazusa DNA Research Institute,  
 Department of Plant Gene Research, 2-6-7 Kazusa-kametarai, Kisarazu,  
 Chiba 292-0818, Japan (E-mail: ssato@kazusa.or.jp,  
 URL: http://www.kazusa.or.jp/, Tel: 81-438-52-3935 (ex.2337),  
 Fax: 81-438-52-3934)  
 FEATURES  
 source location/Qualifiers  
 1..127731  
 /organism="Lotus corniculatus var. japonicus"  
 /mol\_type="genomic DNA"  
 /variant="japonicus"  
 /db\_xref="taxon:34305"  
 /chromosome="4"  
 /clone="UJ79H01"  
 /clone\_id="UJT library"  
 /note="TAC clone:TM0172-synonym: Lotus japonicus"

ORIGIN

Query Match 1.6%; Score 105.4; DB 8; Length 127731;  
 Best Local Similarity 47.0%; Pred. No 1.3e-07;  
 Matches 396; Conservative 0; Mismatches 441; Indels 6; Gaps 2;

QY 4471 CAAATTCAGTATATATGATCCGGATGATTAAGTCAATGATGATCTGAAGAAATT 4530  
 Db 30057 CAAAGAGCTCTGACTGATGATGATCTGATTCATATCATGCAAGAGAACTGGGACAGTT 29998  
 QY 4531 TAGATCAAAAGATGTTTACGAAGAAGTTCCAAATCCCAACGGTGTGAAGCTATATCTAT 4590  
 Db 29997 CAAAGGATATAGATGAGGATGATTTGTTCCAGACTGATGATCAAAATGTTGGGAAC 29938  
 QY 4591 GGGTGGGATACATCTGAAATAATGATTTCTCTCAAGGTGTTGTTGGAATACAGTTG 4650  
 Db 29937 TAAGTGAATATTCAGAAACAAGTCTGATGAAGGTATATGTGACAAGAAACAATCTCG 29878  
 QY 4651 TGTTCATGCGCAACACAAAGAAATTTGGATTTATGACCCCTTTAGTGTATTC 4710  
 Db 29877 TCTAGT---TCTCAGGATCTCTCAATGAAAGATGACCTTTATGAACCTTTTC 29821  
 QY 4711 ACGTGTATATGATCTGATCTATATGATTAATGACATATATAGTGTGATTTAGAAAT 4770  
 Db 29820 TCTGTTGCTGCTCTGATATCTATCAGTTGCTGTTGGAGTACATGTTTGTGAATTT 29761  
 QY 4771 GACAAATCAATTTAGACGTGAGTGGCGGTATCTAAATGCTTATCTACTCTTCAA 4830  
 Db 29760 CAGATTATATCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 29701  
 QY 4831 TCCATTTATGCTTCTCTCCMAATCAGTACCTTTGAAAGAAACCATGTTGTTATTT 4890  
 Db 29700 CTATGTGAAACAACCTTAAGGTTTGTATGATCCAGCTTTCCATATGATGATGAT 29641  
 QY 4891 GAAAGCTTCTGCTATGAGGTTTAAACAGTGGGTTTGGATGTTATCACTATCAAAAG 4950  
 Db 29640 GAAAGAGCTCTGATGCTTAAACAAAGCACTAGACATGATGATGATGATGATGAT 29581  
 QY 4951 ACTATGGAAGACATTTGTTTACTCAAGTTTACAAATGATGATGATGATGATGATGAT 5010  
 Db 29580 ATTCTTATCAACAATGTTATGACAAAGTGGCATGACAAAGCTTTGTTTGAAGA 29521  
 QY 5011 ATATGAGAGGATCAGTATATATTTAGTTTATGTTGATGATATTTCTTATGTTGG 5070  
 Db 29520 AAATGGCGGTGACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 29464  
 QY 5071 AAGTTCAAAAAGTATGATTAATTTTGTGATCAATTTGAGATCACTTTTGAAGTTAA 5130  
 Db 29463 CATGTGCAACCAAGTGGGGAACAATTTGTTGAACAATGATGATGATGATGATGAT 29404  
 QY 5131 AGCTTTGGTGAATATCAATTAATTTCTGTTATGAAATTTGTTAAACCGAATCTGTTA 5190  
 Db 29403 TCTTGTAGTGAATTAATTAATTTCTGTTGACCTTCAAGTAAACAGATGAGATCCTT 29344  
 QY 5191 TATTTTATCTCAAGAAAATTTCTCAGAAAATTTCTTAAGATTTCAACTGATGATCTC 5250  
 Db 29343 ATTCAATCAACAAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 29284  
 QY 5251 ATATGGGAAAAACATACCTGCGATTCGGAATGACAAATATGAAAAGTTGCAATTAATCG 5310  
 Db 29283 TGGTACAAAAGACACCTGCTGCACTCATATCAAGCTGCAAAAGATGAGAAAGGTAC 29224  
 QY 5311 TGA 5313  
 Db 29223 TGA 29221

RESULT 13  
 AP006381 105887 bp DNA linear PLN 22-JUL-2003  
 LOCUS Lotus corniculatus var. japonicus genomic DNA, chromosome 4,  
 clone:UJ79G09, TM0234, complete sequence.  
 ACCESSION AP006381  
 VERSION AP006381.1 GI:31581012  
 KEYWORDS HTG.





\* 11700 29732: contig of 18033 bp in length  
 \* 29733 29832: gap of unknown length  
 \* 29833 129511: contig of 39679 bp in length.

FEATURES  
 Location/Qualifiers

1..129511  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-1036"  
 /clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Query Match 1.6%; Score 103.4; DB 2; Length 129511;  
 Best Local Similarity 46.9%; Pred. No. 2,7e-07;  
 Matches 391; Conservative 0; Mismatches 436; Indels 6; Gaps 2;

4471 CAAATGCACTATATATATCCGGATGATTAAGTCAATGAAGCTAGAGAAATT 4530  
 Db 50154 CAAAGAAAGCCCTACTGATGATGTTGGATTGAAAGCAATGCAATGAAGCTAGATTCATT 50213  
 Qy 4531 TAGATCAAAAGATGTTTACGAGAAGTCCCAATCCCGGTGGAAGCTATATCTAT 4590  
 Db 50214 CAAAGAAAGTGAAGTGAAGTATAGTTCCAGAGCCGAGGTATATATGATGATAGGTAC 50273  
 Qy 4591 GGGTTGGTACATCTGAGAAATTTGATCTCTCAAGGTGTTGCGAAATCAGCTTG 4650  
 Db 50274 CAAAGTATCTATAGAAAAATCTGATGAA--AATGATCTGTAACGAGAAACAGGC 50330  
 Qy 4651 TGTGTCATGCAACAGACAAAGAAAAATGGATTATGACCCCTTTAGTTAGTTTC 4710  
 Db 50331 AAGATTATAGCTTAAGCTACACATATGAAGGTCTGACCTTTATCAGACTTTTGC 50390  
 Qy 4711 ACTGTTATAGATCTTGACTATPAGATTATGACATATATAGTTGTGAATTAGGAAT 4770  
 Db 50391 TCTGTTGCTGATGATGAAATATATAGACTATGCTAGTGTATCATGATACCAAGTT 50450  
 Qy 4771 GACATTTCAACATTAGACGTGCGAGTGGCTATCTAAAGCCCTCTATTTCTCATCAA 4830  
 Db 50451 TAAGTTGTTCAATAGATGTAAGAGAGCTTTTGAATGGTACTGAAAGAGAGAGT 50510  
 Qy 4831 TCCATTATATGCTTCTCTCTAATACGATCCTTGAAGAAAAACCATTTGGTTATT 4890  
 Db 50511 CTATGTTAGCAACCTAAGAGGTTATATGATCTAGCTTCCCATCATGATGATAGATT 50570  
 Qy 4891 GAAAGCTTCTGCTATAGGTTAAAGAGCTGGGTTTGAATGATACACTATCAAG 4950  
 Db 50571 GAAAGAGCTTGTATGCTTAAAGCAGGCACTAGGCTTGGTATGATGACTCA 50630  
 Qy 4951 AGTATGGAAGACATTTGTTTACTCAAGTTTACACATGATGTTTATTTCAATTGA 5010  
 Db 50631 GTTTATTTGTTGTCAGGCTACATGAAGAGGATACATACAGACCTTGT---GTAAA 50687  
 Qy 5011 ATATGAAGAGGATAGTATATATTTAGTTTATGTTATGATATTTCTATAGTTGG 5070  
 Db 50688 AGATACAAAGAGAGAGTATGATCTCATATATATGTTGATACATGTTTTTGG 50747  
 Qy 5071 AAGTCAAAAAGATTATGTAATTTTGGATCAATTGAGATCATTTTGAAGTTAA 5130  
 Db 50748 AATGCAAAAGAGATGTCACAACTTATTCAGCAATGCAATGATGATTTGATGAT 50807  
 Qy 5131 AGTGTGGTGAATATCAATATATCTTTGTTATGATTTGTTAAACCAATGCTTGA 5190  
 Db 50808 CTATGCTGATGATGATGATCTTCTTGGCTTATGTTTAAACAAATGAGATGAT 50867  
 Qy 5191 TATTTATCTCAAGAAAAATTTCTCAAGAAATTTCTTAAGATTTCAACAGATGCTC 5250  
 Db 50868 CTTCATTTCTCAAGATATATATGCAAGACATGATGAAGAAATTTGGCTTGAATATG 50927  
 Qy 5251 ATATGGAAGAAACATACCTCGATTCGAATGACAAATATGAAAAAGTTGCA 5303  
 Db 50928 TGGTATATAGAAAACTCCAACTGACCATTTGTAAGCTGAAAAAGATGAA 50980

RESULT 15  
 AC135396 134544 bp DNA linear HTG 13-DEC-2003  
 LOCUS  
 DEFINITION  
 Medicago truncatula clone mth2-33018, WORKING DRAFT SEQUENCE, 2  
 uncloned pieces.

AC135396  
 AC135396.25 GI:39841085  
 VERSION  
 HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 KEYWORDS  
 Medicago truncatula (barrel medic)  
 SOURCE  
 Medicago truncatula  
 ORGANISM  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustoside I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.

REFERENCE  
 1 (bases 1 to 134544)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,

AUTHORS  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Medicago truncatula BAC Clone mth2-33018

REFERENCE  
 2 (bases 1 to 134544)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission  
 Submitted (14-OCT-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE  
 3 (bases 1 to 134544)  
 Shaull, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
 Cook, D., Kim, D. and Roe, B.A.  
 Direct Submission  
 Submitted (13-DEC-2003) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

COMMENT  
 On Dec 13, 2003 this sequence version replaced gi:39752708.  
 ----- Genome Center  
 The University Of Oklahoma  
 Center code: UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 2 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

\* 1 2202: contig of 2202 bp in length  
 \* 2203 2302: gap of unknown length  
 \* 2303 134544: contig of 13242 bp in length.

FEATURES  
 source  
 1..134544  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-33018"  
 /clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Query Match 1.6%; Score 102.2; DB 2; Length 134544;  
 Best Local Similarity 49.5%; Pred. No. 4.2e-07;  
 Matches 263; Conservative 0; Mismatches 266; Indels 0; Gaps 0;

Qy 4724 CTGTGCACTATATAGTATGCAATATAGTGTGATTTAGGATGACAAATTCACAT 4783  
 Db 30371 CTAGAAAGCTGAGGTATGATACCTATGCTTTGATGAAGATGATGCCACTACCAT 30312  
 Qy 4784 TTAGAGCTGATCGGCGGATGTAATGCTCTATATCTCATTCATTCATTCATTTATGTC 4843  
 Db 30311 TTACATGTAAGTGCACATTTTAAATGATCCGCTGAAGAGGTAGTGTGTGTGCACA 30252  
 Qy 4844 TTCTCTTAATGATGATCTTTGAAGAAAAACCATTTGTTGTTAAGACGTTCTGTC 4903

Db	30251	CCACAGGTTTGAAATTCGTAGAGAAAGAAAATGTGGTATATACTGACATAAGACATA	30122
Qy	4904	TATGGGTTAAACACGTGGGTTTGAAATGCTATCACATATCAAAAAGATATGGAGAC	4963
Db	30191	TATGAGATTAACACGGCTCCTAGAGCATGTGCACAGAAAATGATCAAGTTTAAATTCAA	30132
Qy	4964	ATTGGCTTTACAGGTTTACACAGATGCTTTATTTCATCTTGAAATATGAAAGGGA	5023
Db	30131	ATTAGGCTTTAAAAAGTGTCAAGTTGAGTTTGGGTATATGTTTCAGAAAGCCTAGTATGA	30072
Qy	5024	TCAGTATATATATTAGGTTAATATGTTGATGATATTTCTTATGTTGGAAGTTCACAAAA	5083
Db	30071	GGAGCTGTATAAATTTGTCTGTATGTTGATGACTTACTATCATCTGTAGCTCAACATCA	30012
Qy	5084	GTTATGTATATATTTTGTGATCAATTGAGAGATCATTTTGAAGTTAAAGTTTGGTGAA	5143
Db	30011	GAGATTTAAGAAAGGTTAAAGAAACCTGAAGCTGAGTTTGAGATGACTGATCTTGGTGAG	29952
Qy	5144	ATATCAAAATTAATCTTGGTATGTGAATTTTCGTAAAAACCGAATCTGGTATATTTATCTCAA	5203
Db	29951	CTTTCATCTTCTTTAGGAATGGAATTTTGAAAAATGAAAAGTTGGGATAGTGATGATCA	29892
Qy	5204	GAAAATTTTCCAGAAATTAATCTTAAGGATTTTCAAACTGATGATCCATAT	5254
Db	29891	AAAAAATTCATTTGGTGAAATTCCTTGATAAATTTTGAAATGAAGAGCTGCAT	29841

Search completed: April 11, 2004, 18:29:14  
Job time : 15982 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 07:57:56 ; Search time 1502 Seconds

(without alignments)  
18175.047 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426  
Sequence: 1 tgcgtggttcgtgcacattt.....agaaagttacatccatca 6426

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: N\_Geneseq\_29Jan04:\*

1: geneseqn1980s:\*\n2: geneseqn1990s:\*\n3: geneseqn2000s:\*\n4: geneseqn2001as:\*\n5: geneseqn2001bs:\*\n6: geneseqn2002s:\*\n7: geneseqn2003as:\*\n8: geneseqn2003bs:\*\n9: geneseqn2003cs:\*\n10: geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6424.4	100.0	6426	3	AAAS7920 Candida a
2	392	6.1	392	3	AAAS7936
3	388.8	6.1	392	3	AAAS7928
4	388.8	6.1	392	3	AAAS7931
5	388.8	6.1	392	3	AAAS7923
6	388.8	6.1	392	3	AAAS7932
7	388.8	6.1	392	3	AAAS7927
8	387.2	6.0	392	3	AAAS7925
9	387.2	6.0	392	3	AAAS7933
10	387.2	6.0	392	3	AAAS7929
11	385.6	6.0	392	3	AAAS7930
12	384	6.0	392	3	AAAS7934
13	384	6.0	392	3	AAAS7935
14	382.4	6.0	392	3	AAAS7924
15	380.8	5.9	392	3	AAAS7926
16	328	5.1	1385	3	AAAS7989
17	280	4.4	1249	3	AAAS7947
18	280	4.4	1483	3	AAAS7990
19	280	4.4	5611	3	AAAS7948
20	275.2	4.1	879	3	AAAS7991
21	264.4	4.3	454	3	AAAS7967
22	255.6	4.0	974	3	AAAS7992
23	252.8	3.9	1308	3	AAAS7949

24	249.8	3.9	469	3	AAAS7994	AAAS7994 469 bp Ca
25	224.8	3.5	9850	3	AAAS8018	AAAS8018 9850 bp C
26	224.6	3.5	3604	3	AAAS7946	AAAS7946 3604 bp C
27	190	3.0	690	3	AAAS7951	AAAS7951 690 bp Ca
28	170.4	2.7	1912	3	AAAS7952	AAAS7952 1912 bp C
29	166	2.6	1688	4	AAAS9873	AAAS9873 C albican
30	164	2.6	1672	3	AAAS7950	AAAS7950 1672 bp C
31	130	2.0	138	3	AAAS7937	AAAS7937 Candida a
32	104.2	1.6	770	3	AAAS7964	AAAS7964 770 bp Ca
33	104.2	1.6	951	3	AAAS8017	AAAS8017 951 bp Ca
34	100.6	1.6	1348	3	AAAS7983	AAAS7983 1348 bp C
35	93.2	1.5	1340	3	AAAS7940	AAAS7940 1340 bp C
36	91	1.4	91	3	AAAS7922	AAAS7922 Candida a
37	86.4	1.3	12810	3	AAAS5272	AAAS5272 Zmmta me
38	83.8	1.3	13340	2	AAAS3522	AAAS3522 O. longis
39	72.2	1.1	96988	3	AAAS2290	AAAS2290 BAC conta
40	71.4	1.1	3780	9	ADCO8114	ADCO8114 Rice DNA
41	71.2	1.1	6486	6	ABOG7050	ABOG7050 Human ang
42	70.6	1.1	110000	3	AAAS2305_06	Continuation (7 of
43	70.2	1.1	1298	3	AAAS70212	AAAS70212 plasmodiu
44	68.2	1.1	7306	6	ABIS3637	ABIS3637 Human imm
45	67.2	1.0	1083	5	AAAS76745	AAAS76745 DNA encod

## ALIGNMENTS

RESULT 1	AAAS7920	AAAS7920 standard; DNA; 6426 BP.
XX	AAAS7920	
AC	AAAS7920	
XX	15-SEP-2003 (revised)	
DT	11-OCT-2000 (first entry)	
XX	Candida albicans pcal retrotransposon.	
DE	Candida albicans pcal retrotransposon.	
XX	Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;	
KW	gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;	
KW	reverse transcriptase; RNaseH; pseudoknot; readthrough translation;	
KW	stop codon suppression; gene delivery; gene therapy vector;	
KW	genetic vaccine composition; immunogenic; transgenic animal; ds.	
XX		
OS	Candida albicans; strain HOG1042.	
XX		
FT	Key	Location/Qualifiers
FT	LTR	1..280
FT	repeat_unit	/*tag= a
FT	repeat_unit	1..6
FT	TATA_signal	/rpt_type= INVERTED
FT	TATA_signal	/note= "imperfect 6 bp repeat"
FT	TATA_signal	120..125
FT	TATA_signal	/*tag= c
FT	TATA_signal	/standard_name= "TATA box"
FT	TATA_signal	185..190
FT	TATA_signal	/*tag= d
FT	TATA_signal	/standard_name= "TATA box"
FT	TATA_signal	201..206
FT	polya_signal	/*tag= e
FT	polya_signal	275..280
FT	repeat_unit	/*tag= f
FT	repeat_unit	/rpt_type= INVERTED
FT	repeat_unit	/note= "imperfect 6 bp repeat"
FT	repeat_unit	281..291
FT	primer_bind	/*tag= g
FT	primer_bind	/note= "This site in the corresponding RNA is a primer
FT	primer_bind	binding site for reverse transcriptase-mediated minus-
FT	primer_bind	strand DNA synthesis"
FT	primer_bind	498..6103
FT	primer_bind	/*tag= 1
FT	primer_bind	/product= "Gag (group antigen)-pol (polyprotein)"

```

FT readthrough protein"
FT /note= "The gag-pol readthrough protein is the result of
FT stop codon suppression mediated by pseudoknot formation
FT in the mRNA"
FT
FT /codon= (seq: "ctg", aa: Ser)
FT /note= "12 serine residues in the gag-pol readthrough
FT protein are encoded by this non-standard Ser codon"
FT /transl_except= (pos: 1167..1372, aa: Gln)
FT 498..1372
FT
FT /tag= h
FT /product= "Gag (group antigen) protein"
FT /codon= (seq: "ctg", aa: Ser)
FT /note= "5 serine residues in the gag protein are encoded
FT by this non-standard Ser codon"
FT 1373..6100
FT
FT /tag= x
FT /product= "Pol (polypotein), comprising aspartate
FT /product= "Integrase, reverse transcriptase (RT) and
FT RNaseH"
FT /codon= (seq: "ctg", aa: Ser)
FT /note= "7 serine residues in the pol protein are encoded
FT by this non-standard Ser codon"
FT 1381..1455
FT
FT /tag= j
FT /note= "The RNA corresponding to this region forms a
FT pseudoknot, allowing gag ORF stop codon suppression and
FT translation of the gag-pol readthrough protein"
FT 3455..3465
FT
FT /tag= l
FT /note= "Polypurine tract 2 (PPT2)"
FT complement(3455..3465)
FT
FT /tag= m
FT /note= "This site in the corresponding RNA is a primer
FT binding site for reverse transcriptase-mediated plus-
FT strand DNA synthesis"
FT 6136..6146
FT
FT /tag= n
FT /note= "Polypurine tract 1 (PPT1)"
FT complement(6136..6146)
FT
FT /tag= o
FT /note= "This site in the corresponding RNA is a primer
FT binding site for reverse transcriptase-mediated plus-
FT strand DNA synthesis"
FT 6147..6426
FT
FT /tag= p
FT 6147..6152
FT
FT /tag= q
FT /rpt_type= INVERTED
FT /note= "Imperfect 6 bp repeat"
FT 6266..6271
FT
FT /tag= r
FT /standard_name= "TATA box"
FT 6331..6336
FT
FT /tag= s
FT /standard_name= "TATA box"
FT 6347..6351
FT
FT /tag= t
FT 6421..6426
FT
FT /tag= u
FT /rpt_type= INVERTED
FT /note= "Imperfect 6 bp repeat"

```

PI Luyten WHMGL, De Backer MD, Nelliesen BDM, Foulter RTM;  
 XX  
 DR WPI: 2000-365640/31.  
 DR P-PSDB: AAB03126, AAB03127, AAB03128.  
 XX  
 XX Novel retrovirusposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 XX Claim 12; Fig 2B; 204pp; English.  
 XX

The invention relates to novel retrotransposons from the yeast *Candida albicans* which have a copy number of 40-150, preferably 50-100 copies per genome. In particular, the invention relates to the novel *C. albicans* *TY1* copia retrotransposon *pGal* (AA55/920), and to the integrated form of this retrotransposon, designated *TGα2*, and to the novel *C. albicans* retrotransposons 1-28. *pGal* was initially isolated from *C. albicans* HO31042 and has a copy number of 50-100 copies per cell. It comprises identical 280 bp long terminal repeats (LTRs) and two open reading frames (ORFs). The first ORF encodes a gag (group antigen) protein, and the second ORF encodes a polypeptide (pol) consisting of an aspartate protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and pol ORFs of *pGal* are in the same reading frame, separated only by a termination codon (TGA). Translation of the pol ORF occurs through the occasional readthrough suppression of the stop codon, which is mediated by the formation of a pseudoknot within the gag-pol mRNA. The retrotransposons of the invention can be used as vectors for *in vitro* or *in vivo* transformation and expression. They can thus be used for the delivery and expression of a therapeutic, immunological or immunogenic molecule (e.g., an antigen) and may also be used for eliciting an immunological response in a host organism. They are therefore useful in immunologic vaccine compositions and for gene therapy, particularly where the use of retroviral vectors is unsafe or undesirable. Additionally, the retrotransposons may be used to generate transgenic animals, to detect the presence of *Candida* in a sample, to detect and disrupt genes, and to assign functions to nucleotide sequences. The present sequence represents the *C. albicans* *pGal* retrotransposon. (Updated on 15-SEP-2003 to standardise OS field)

Query Match	100.0%	Score 6424.4	DB 3	Length 6426
Best Local Similarity	100.0%	Pred. No. 0		
Matches 6425; Conservative	0	Mismatches	1	Indels 0; Gaps 0

QY	1	TGTTGGTTGTGACACATATTTGTGTCAGAAATCGATGCAATGAAAATGATGCTATTTTGA	60
Db	1	TGTTGGTTGTGACACATATTTGTGTCAGAAATCGATGCAATGAAAATGATGCTATTTTGA	60
QY	61	GAATGGAATATTTTTCATCACAACATCAGTGATCAGAACTAACTATATTGTAGT	120
Db	61	GAATGGAATATTTTTCATCACAACATCAGTGATCAGAACTAACTATATTGTAGT	120
QY	121	ATAAATAGGGTATGAATAACCAACATCCAGATATCAACGAGTAGAAGGAGGAGTT	180
Db	121	ATAAATAGGGTATGAATAACCAACATCCAGATATCAACGAGTAGAAGGAGGAGTT	180
QY	181	TCAAATATATCTTTGATGATTAATTAACCTGGTCTAATCACAATACACATTAAGACGCT	240
Db	181	TCAAATATATCTTTGATGATTAATTAACCTGGTCTAATCACAATACACATTAAGACGCT	240
QY	241	ACAAGCTCATCTCAGGTAAAGAAAGTTATATTCATCAGATTAGAAGTCAGATGAT	300
Db	241	ACAAGCTCATCTCAGGTAAAGAAAGTTATATTCATCAGATTAGAAGTCAGATGAT	300
QY	301	AATCATTTCGCCAATTAAGCGTTGTATAAATTAAGTCTTCAGATTGTATTTATGATT	360
Db	301	AATCATTTCGCCAATTAAGCGTTGTATAAATTAAGTCTTCAGATTGTATTTATGATT	360
QY	361	GATAGTTTGAAGTTGAAGGTACAGAAATTTACAGAGATGAGTTCCGCAAAAGATGATGA	420
Db	361	GATAGTTTGAAGTTGAAGGTACAGAAATTTACAGAGATGAGTTCCGCAAAAGATGATGA	420

```
QY 421 TAACGAGGAAAGTGCATGAGAAAGTGGATCAAGCTAATGCTATTAGTAGTGATGA 480
DB 421 TAACGAGGAAAGTGCATGAGAAAGTGGATCAAGCTAATGCTATTAGTAGTGATGA 480
QY 481 ACATATCAAGGCTAGATTCATATGCTTTTCAATAATTAAATGAAGCTTAAGTGGC 540
DB 481 ACATATCAAGGCTAGATTCATATGCTTTTCAATAATTAAATGAAGCTTAAGTGGC 540
QY 541 CGTCGATATCAGAAAGGCTGATTAATGAGTAAGAAATTAATTTCCAGTGGC 600
DB 541 CGTCGATATCAGAAAGGCTGATTAATGAGTAAGAAATTAATTTCCAGTGGC 600
QY 601 TTACCCCGATGTTTGGAAATTTTTCCTGACTAATACTTAAAGATAATTCAGAGTTAA 660
DB 601 TTACCCCGATGTTTGGAAATTTTTCCTGACTAATACTTAAAGATAATTCAGAGTTAA 660
QY 661 AAAGGTAGAGGATTTATTTTACGTTGCTGTTTACAAATGCTTACAGTCCATTTT 720
DB 661 AAAGGTAGAGGATTTATTTTACGTTGCTGTTTACAAATGCTTACAGTCCATTTT 720
QY 721 TGATAGGTTCAAGATTGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGAGCAA 780
DB 721 TGATAGGTTCAAGATTGATCATGATTTCTAAGCTACCAAGCACTTGCAAAAGAGCAA 780
QY 781 CTTAATCAAGGCTGCTTATGATGCTGTTTCTAATCTTAAAGTTATACCATTTACTAGTA 840
DB 781 CTTAATCAAGGCTGCTTATGATGCTGTTTCTAATCTTAAAGTTATACCATTTACTAGTA 840
QY 841 GATCTTGCTGAGAGTTTGTAAAGTGAACATGAGTTAGTGGTTCCTAATCTTCCATA 900
DB 841 GATCTTGCTGAGAGTTTGTAAAGTGAACATGAGTTAGTGGTTCCTAATCTTCCATA 900
QY 901 TTTGCTGAGGTTGAGAAAGAAACCTTGAGAAATCTTACCAACACTTCAAAAGTTTTCGA 960
DB 901 TTTGCTGAGGTTGAGAAAGAAACCTTGAGAAATCTTACCAACACTTCAAAAGTTTTCGA 960
QY 961 TGAGATGTCGCTGATCTTCCAAATCTGATAGGTCGAAGCTTGTACTTCAATCATGTGAA 1020
DB 961 TGAGATGTCGCTGATCTTCCAAATCTGATAGGTCGAAGCTTGTACTTCAATCATGTGAA 1020
QY 1021 GAATCAGAGGCTTTAAGTTGTTTGTGAATATTCATGCTCATACTAATCAAGTGGAT 1080
DB 1021 GAATCAGAGGCTTTAAGTTGTTTGTGAATATTCATGCTCATACTAATCAAGTGGAT 1080
QY 1081 TCAAGCTGACAAATGATACATCTGATCTCCCAAGTTCCTTACCATAGCTGAGAAATGTG 1140
DB 1081 TCAAGCTGACAAATGATACATCTGATCTCCCAAGTTCCTTACCATAGCTGAGAAATGTG 1140
QY 1141 TGATATCTGCTGATTTATGCTAGATGCTGATCTCCCAAGTTCCTTACCATAGCTGAG 1200
DB 1141 TGATATCTGCTGATTTATGCTAGATGCTGATCTCCCAAGTTCCTTACCATAGCTGAG 1200
QY 1201 TATTTGTTAGTTTACAGACACAGAGAAACCAAAAGAAACCAAGAGAACTCACTGGA 1260
DB 1201 TATTTGTTAGTTTACAGACACAGAGAAACCAAAAGAAACCAAGAGAACTCACTGGA 1260
QY 1261 ACAATCTCAAAAGAAAGAACTGTAATCAAGAAAGAAATTAAGAAACATCCAAATGAGA 1320
DB 1261 ACAATCTCAAAAGAAAGAACTGTAATCAAGAAAGAAATTAAGAAACATCCAAATGAGA 1320
QY 1321 TAAAGATAAGGTTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1380
DB 1321 TAAAGATAAGGTTGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAG 1380
QY 1381 TGCTGCTCTTATTAATGCTGTAATGAATATACATAATGACAGAAAAACCAAGTTTCCAGT 1440
DB 1381 TGCTGCTCTTATTAATGCTGTAATGAATATACATAATGACAGAAAAACCAAGTTTCCAGT 1440
QY 1441 AGAAATATCTCATTTCTCTTAATGCTTTTGAAGCTAATGTAATTTTAAAGGTTTAAAGTT 1500
DB 1441 AGAAATATCTCATTTCTCTTAATGCTTTTGAAGCTAATGTAATTTTAAAGGTTTAAAGTT 1500
QY 1501 TAAAGATATCTAGTATGATAGTACGTCGCCAATATCTGTTGAAACAAATTAAGATAT 1560
DB 1501 TAAAGATATCTAGTATGATAGTACGTCGCCAATATCTGTTGAAACAAATTAAGATAT 1560

QY 1501 TAACAGATATCTAGTATGATAGTACGTCGCCAATATCTGTTGTAACAAATTAAGATAT 1560
DB 1501 TAACAGATATCTAGTATGATAGTACGTCGCCAATATCTGTTGTAACAAATTAAGATAT 1560
QY 1561 ATTGCTGAATGTTAAGACGCAACCAATTAAGTTTCTGTTGCTGATGAGTGCTTACATTA 1620
DB 1561 ATTGCTGAATGTTAAGACGCAACCAATTAAGTTTCTGTTGCTGATGAGTGCTTACATTA 1620
QY 1621 AGCAGATGATATGCTGATTAATTAATCAAGAGCTGATTTGCTGATTAAGTAAGAA 1680
DB 1621 AGCAGATGATATGCTGATTAATTAATCAAGAGCTGATTTGCTGATTAAGTAAGAA 1680
QY 1681 TACATGATATTAACGAAAGGTTCTTTAATCTTGTGAGTTTGAACCAATTAAGTAAG 1740
DB 1681 TACATGATATTAACGAAAGGTTCTTTAATCTTGTGAGTTTGAACCAATTAAGTAAG 1740
QY 1741 AGGATTTAATGCTTATTAATTAAGAAATCAGATGTTGATTTAATCAAAATGCTGCTC 1800
DB 1741 AGGATTTAATGCTTATTAATTAAGAAATCAGATGTTGATTTAATCAAAATGCTGCTC 1800
QY 1801 TACTATTAATGCTTCAAGAGAAATGCTGATCTTATATAGGTCCTCAATTCAGTGA 1860
DB 1801 TACTATTAATGCTTCAAGAGAAATGCTGATCTTATATAGGTCCTCAATTCAGTGA 1860
QY 1861 AGAATCTTTAGATGTTGATTTTATGATAGTGTGCGAGATAGTTGCTCCAAATGCTTA 1920
DB 1861 AGAATCTTTAGATGTTGATTTTATGATAGTGTGCGAGATAGTTGCTCCAAATGCTTA 1920
QY 1921 CCAAGATGACAAAGATTAATCAAGATGATGATGAATGTCAGAAATCAAGAAATGATTA 1980
DB 1921 CCAAGATGACAAAGATTAATCAAGATGATGATGAATGTCAGAAATCAAGAAATGATTA 1980
QY 1981 TAGTCTCGAGCATTAATAATCTTTTGAACGAGTTGATGTTTGAATGTTGAATTTTC 2040
DB 1981 TAGTCTCGAGCATTAATAATCTTTTGAACGAGTTGATGTTTGAATGTTGAATTTTC 2040
QY 2041 CCCATATGAGGTTGAACAAATGCTTACCAACTGAGATTAAGAAAGATTTATTAATTTCA 2100
DB 2041 CCCATATGAGGTTGAACAAATGCTTACCAACTGAGATTAAGAAAGATTTATTAATTTCA 2100
QY 2101 TTTGATGTCAAATCATATGTCATTTGAGAAATCTTGTGTTACAAAATACAGAGGTC 2160
DB 2101 TTTGATGTCAAATCATATGTCATTTGAGAAATCTTGTGTTACAAAATACAGAGGTC 2160
QY 2161 CGTACTTCAACTTCAAAAGAGGCTTCAAAAGATGCTGATGTTGAAGTATGCTAT 2220
DB 2161 CGTACTTCAACTTCAAAAGAGGCTTCAAAAGATGCTGATGTTGAAGTATGCTAT 2220
QY 2221 ATCGAATGCGCAACAGAGAAATGCAATCATCATCTTCAAGAAAGAAAGCTCGAGAGACA 2280
DB 2221 ATCGAATGCGCAACAGAGAAATGCAATCATCATCTTCAAGAAAGAAAGCTCGAGAGACA 2280
QY 2281 TGAAGACTTCAATGCTGATCTCTGCTGATTTAGTCCGAAATTAACAAGTGTATTT 2340
DB 2281 TGAAGACTTCAATGCTGATCTCTGCTGATTTAGTCCGAAATTAACAAGTGTATTT 2340
QY 2341 AACGCTGTTATGATGAGACATACGAGTTTACATTTGAAGAAATTAATTAACAGAGAA 2400
DB 2341 AACGCTGTTATGATGAGACATACGAGTTTACATTTGAAGAAATTAATTAACAGAGAA 2400
QY 2401 GGTAAAGATCTCTTAATTAACAGATTAAGATCTGGAATTAATGCTTTAAGATTAAGT 2460
DB 2401 GGTAAAGATCTCTTAATTAACAGATTAAGATCTGGAATTAATGCTTTAAGATTAAGT 2460
QY 2461 GGCATCTTCAAGAGATTAATGCTGATCTGATCTCCAGAACTTCTGATTTAGCTAGTT 2520
DB 2461 GGCATCTTCAAGAGATTAATGCTGATCTGATCTCCAGAACTTCTGATTTAGCTAGTT 2520
QY 2521 CGGATTTTGAAGGAGATTAAGCGGATATGCTGAGCTTAATGCTGCGCGAGGT 2580
DB 2521 CGGATTTTGAAGGAGATTAAGCGGATATGCTGAGCTTAATGCTGCGCGAGGT 2580
QY 2581 TGTTAATAATGATTTTACAGATTAAGAGATGCTTGAACCTGTCACCAAT 2640
DB 2581 TGTTAATAATGATTTTACAGATTAAGAGATGCTTGAACCTGTCACCAAT 2640
```

Db 2581 TGTAAATTAATGATTTTACACAGATTTACAGAGATGTTGTGCACTTGTCCACAAAT 2640  
Qy 2641 ACTCAAGTTGATTTATATATGATTTCAATATTCATATACATATGATCAACCACTCCACG 2700  
Db 2641 ACTCAAGTTGATTTATATATGATTTCAATATTCATATACATATGATCAACCACTCCACG 2700  
Qy 2701 TCGTTCACTCAAGGACAAACCCCTTATGTTGCTATTAATCAATTAAGTAGGGAATTT 2760  
Db 2701 TCGTTCACTCAAGGACAAACCCCTTATGTTGCTATTAATCAATTAAGTAGGGAATTT 2760  
Qy 2761 CTACCGGTTTCTTTTGGCATGCAATGTGTGCTATCACTTTAGTAATCCCATCGAAAAGA 2820  
Db 2761 CTACCGGTTTCTTTTGGCATGCAATGTGTGCTATCACTTTAGTAATCCCATCGAAAAGA 2820  
Qy 2821 CCGTTACGAGTTCATCACTTAAGAGAGCTCTTCATCGATCATGGGCTGTGATTGG 2880  
Db 2821 CCGTTACGAGTTCATCACTTAAGAGAGAGCTCTTCATCGATCATGGGCTGTGATTGG 2880  
Qy 2881 CTACGCTAGCGATTGTTTACTATTAACGTGTCTTAAATAATGCGGTGTGATTTAT 2940  
Db 2881 CTACGCTAGCGATTGTTTACTATTAACGTGTCTTAAATAATGCGGTGTGATTTAT 2940  
Qy 2941 CCTTACGCTTATGTCGTATATGCGAGGTATGAGGTTATTAATCTCTCAAAA 3000  
Db 2941 CCTTACGCTTATGTCGTATATGCGAGGTATGAGGTTATTAATCTCTCAAAA 3000  
Qy 3001 CTATTCACATACCTATATGTCACATGTTCTATGCTGAGGTTATTCAGGAGGCA 3060  
Db 3001 CTATTCACATACCTATATGTCACATGTTCTATGCTGAGGTTATTCAGGAGGCA 3060  
Qy 3061 ACTGGGCGCTCAGTACGAGGTACGCGAACAATGTGGAAGTGAATGCAATACAA 3120  
Db 3061 ACTGGGCGCTCAGTACGAGGTACGCGAACAATGTGGAAGTGAATGCAATACAA 3120  
Qy 3121 TGACGTGATGACATGCCCCAAGAGTCATATTCAGTTCAGCCAGCATGTTTACTTAA 3180  
Db 3121 TGACGTGATGACATGCCCCAAGAGTCATATTCAGTTCAGCCAGCATGTTTACTTAA 3180  
Qy 3181 TACGGGTACAGTTCATACGAATATGTTAAATGATGATCCAGTACAGATTAACATTA 3240  
Db 3181 TACGGGTACAGTTCATACGAATATGTTAAATGATGATCCAGTACAGATTAACATTA 3240  
Qy 3241 GAATCCGATGATTTTCTTAACTCTTCACTCACTAAGTGAAGTCAACGATATGATTC 3300  
Db 3241 GAATCCGATGATTTTCTTAACTCTTCACTCACTAAGTGAAGTCAACGATATGATTC 3300  
Qy 3301 CGAAGTAAATCGAGTGAATTCCTAAACCCAGTCTCCAGAGCTTAAACCTGGGATTA 3360  
Db 3301 CGAAGTAAATCGAGTGAATTCCTAAACCCAGTCTCCAGAGCTTAAACCTGGGATTA 3360  
Qy 3361 TCCGGTGTAAACCTCTCACTTGTGATCGAGAGCTTCACTATTAAGGAACTTAAAGA 3420  
Db 3361 TCCGGTGTAAACCTCTCACTTGTGATCGAGAGCTTCACTATTAAGGAACTTAAAGA 3420  
Qy 3421 GCTTATTAACAACACAAAGAGAGCCCTTCATCCAGGGAGGACCAATAAGCCT 3480  
Db 3421 GCTTATTAACAACACAAAGAGAGCCCTTCATCCAGGGAGGACCAATAAGCCT 3480  
Qy 3481 GGAATCTACTGCTCAGTTGATGATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3540  
Db 3481 GGAATCTACTGCTCAGTTGATGATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3540  
Qy 3541 GGAATCTACTGCTCAGTTGATGATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3600  
Db 3541 GGAATCTACTGCTCAGTTGATGATCAACCAACCCAGAGCTGTCTCCCGTTGGA 3600  
Qy 3601 AGAATATGGCATATCTTCTGACTACCGAGAACTATGCTGAAGTGAACAGCAAAATCC 3660  
Db 3601 AGAATATGGCATATCTTCTGACTACCGAGAACTATGCTGAAGTGAACAGCAAAATCC 3660  
Qy 3661 CTCGTTACTGCTAATGAAGTGGGTACTGAAAAATAGATAGGGAGAAATATTTTC 3720  
Db 3661 CTCGTTACTGCTAATGAAGTGGGTACTGAAAAATAGATAGGGAGAAATATTTTC 3720

Qy 3721 ATTTCCGGGGGTGATGATGATTCGTGCTGATCACTCAATGTTGACATTTAATGT 3780  
Db 3721 ATTTCCGGGGGTGATGATGATTCGTGCTGATCACTCAATGTTGACATTTAATGT 3780  
Qy 3781 TGAACAGAGATGCTGTAAACAGTCCATTCAGAGAGAGTTCCTCAAGGGAGAGAT 3840  
Db 3781 TGAACAGAGATGCTGTAAACAGTCCATTCAGAGAGAGTTCCTCAAGGGAGAGAT 3840  
Qy 3841 ACTAATGAACAACTGATATAGTTGATGATGATGATGATGATGATGATGATGATGAT 3900  
Db 3841 ACTAATGAACAACTGATATAGTTGATGATGATGATGATGATGATGATGATGATGAT 3900  
Qy 3901 CTCTCTTATTAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Db 3901 CTCTCTTATTAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3960  
Qy 3961 CAATCAACAGATCCGATTCATTCAGAAATGAATATCACAGTATTAATGAGA 4020  
Db 3961 CAATCAACAGATCCGATTCATTCAGAAATGAATATCACAGTATTAATGAGA 4020  
Qy 4021 AAATACGAATATATCCAAAAACATTTGAAGTATCTTCTGCTGATTAAGATTTGATGA 4080  
Db 4021 AAATACGAATATATCCAAAAACATTTGAAGTATCTTCTGCTGATTAAGATTTGATGA 4080  
Qy 4081 ATTTGAACGATATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
Db 4081 ATTTGAACGATATATGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4140  
Qy 4141 AGCTAATCCACTCAGATGAAATATGATGTTCAATGATGATGATGATGATGATGATGAT 4200  
Db 4141 AGCTAATCCACTCAGATGAAATATGATGTTCAATGATGATGATGATGATGATGATGAT 4200  
Qy 4201 TATAGCATGATGACAGCAAGAAATTAACAATTTGAGAAAGTGAAGTGAAGTGAAGT 4260  
Db 4201 TATAGCATGATGACAGCAAGAAATTAACAATTTGAGAAAGTGAAGTGAAGTGAAGT 4260  
Qy 4261 TCTGTGTATTAACATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320  
Db 4261 TCTGTGTATTAACATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320  
Qy 4321 TACTGTGTAACCAAGAGATGAAACCAAGAAAGAGAGTGTCTATCACTGTGAAT 4380  
Db 4321 TACTGTGTAACCAAGAGATGAAACCAAGAAAGAGAGTGTCTATCACTGTGAAT 4380  
Qy 4381 AAACAAAGATGCAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4440  
Db 4381 AAACAAAGATGCAATCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4440  
Qy 4441 CCCAAGAGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500  
Db 4441 CCCAAGAGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4500  
Qy 4501 AAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
Db 4501 AAAGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560  
Qy 4561 AATTCACACGGTGTGAAGCCATATCTATGAGTGTGGTATACATCTGAGAAATGATTC 4620  
Db 4561 AATTCACACGGTGTGAAGCCATATCTATGAGTGTGGTATACATCTGAGAAATGATTC 4620  
Qy 4621 TCTCAAGGTGTTTTCGAAATTCAGTGTGTCTCAATGAGCAACACAAAGAGAA 4680  
Db 4621 TCTCAAGGTGTTTTCGAAATTCAGTGTGTCTCAATGAGCAACACAAAGAGAA 4680  
Qy 4681 ATTTGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
Db 4681 ATTTGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4740  
Qy 4741 ATTTGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800  
Db 4741 ATTTGATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4800



QY 4801 GATCTAAATGCTCTATTACTCATTCAAATCCAAATTTATGTCTTCTCTCTAAATCAGT 4860  
 DB 4801 GATCTAAATGCTCTATTACTCATTCAAATCCAAATTTATGTCTTCTCTCTAAATCAGT 4860  
 QY 4861 ACCITTTGAAGAAAACCATTTGTTGTTATTTGAAAGTTCTGTCTATGGGTTAAAAAGTGC 4920  
 DB 4861 ACCITTTGAAGAAAACCATTTGTTGTTATTTGAAAGTTCTGTCTATGGGTTAAAAAGTGC 4920  
 QY 4921 GGGTTTGAATGCTATCACACTATCAAAAAGATTTGGAAGACATTTGTTTACTCAAGT 4980  
 DB 4921 GGGTTTGAATGCTATCACACTATCAAAAAGATTTGGAAGACATTTGTTTACTCAAGT 4980  
 QY 4981 TTTACCAATGATGGTTTATTTTCACTGAATGATGAAGGGAATGAGTAATATTTTGG 5040  
 DB 4981 TTTACCAATGATGGTTTATTTTCACTGAATGATGAAGGGAATGAGTAATATTTTGG 5040  
 QY 5041 TTTAATATGTTGATGATTTCTTATGTTGGAAGTTCACAAAAAGTTATTTGATTTTGT 5100  
 DB 5041 TTTAATATGTTGATGATTTCTTATGTTGGAAGTTCACAAAAAGTTATTTGATTTTGT 5100  
 QY 5101 GGATCAATGAGATGATTTTGAAGTTTAAAGTTTGGTGAATATCAAAATTTATCTTGG 5160  
 DB 5101 GGATCAATGAGATGATTTTGAAGTTTAAAGTTTGGTGAATATCAAAATTTATCTTGG 5160  
 QY 5161 TATTGAATTTGTTAAACCGAATCTGTTTATTTATCTCAAGAAAATTTCTCAAGAA 5220  
 DB 5161 TATTGAATTTGTTAAACCGAATCTGTTTATTTATCTCAAGAAAATTTCTCAAGAA 5220  
 QY 5221 ATTACTTAAGATTTTCAAACTTGATGACTCATATGGGAAAAACATACCTTGATTCGAA 5280  
 DB 5221 ATTACTTAAGATTTTCAAACTTGATGACTCATATGGGAAAAACATACCTTGATTCGAA 5280  
 QY 5281 TGACAAATATGAAAGTTGCGAATATTCGTAAGAAAACGTTATCCAGAGAAATGATTTTGA 5340  
 DB 5281 TGACAAATATGAAAGTTGCGAATATTCGTAAGAAAACGTTATCCAGAGAAATGATTTTGA 5340  
 QY 5341 AAAGGTTCCGAATGAGACATTTGCTGACCTGATGCTAAAAAAGTATACCAAAAGTGTGT 5400  
 DB 5341 AAAGGTTCCGAATGAGACATTTGCTGACCTGATGCTAAAAAAGTATACCAAAAGTGTGT 5400  
 QY 5401 TGGCCTGCTTTTATGGGCGTCCAGAAACACACGTCGCAATATTCGGTCTGATGAAATTC 5460  
 DB 5401 TGGCCTGCTTTTATGGGCGTCCAGAAACACACGTCGCAATATTCGGTCTGATGAAATTC 5460  
 QY 5461 GTTGGGTTCTAAATCTGCAAAATCCAAATGCTCATGATATGAGAAATTTATTTGCT 5520  
 DB 5461 GTTGGGTTCTAAATCTGCAAAATCCAAATGCTCATGATATGAGAAATTTATTTGCT 5520  
 QY 5521 TAGGTAATTCAAAAATACATGAGGATATCATGATGAGTACAAAAGAAACAGATTTGAAAT 5580  
 DB 5521 TAGGTAATTCAAAAATACATGAGGATATCATGATGAGTACAAAAGAAACAGATTTGAAAT 5580  
 QY 5581 ACCACCAAAATCATTTGTTATGCAATGTTTCAAGTATGCTGATTTGACACGAGATTTGA 5640  
 DB 5581 ACCACCAAAATCATTTGTTATGCAATGTTTCAAGTATGCTGATTTGACACGAGATTTGA 5640  
 QY 5641 TAGAAATCTATGAGTGAACCTTTGATTTATGTAATGGAATTTGTTGCAATGGGCGAC 5700  
 DB 5641 TAGAAATCTATGAGTGAACCTTTGATTTATGTAATGGAATTTGTTGCAATGGGCGAC 5700  
 QY 5701 CAAAAAACAAAGGATGATGACAAAGGTCAGAGCTTTGGAATGTTGGTCTTAATTA 5760  
 DB 5701 CAAAAAACAAAGGATGATGACAAAGGTCAGAGCTTTGGAATGTTGGTCTTAATTA 5760  
 QY 5761 TACAATGTTGAAGCTATGCAAAATAAAAACCATTTAATGATTTGGGTTTGAAGTAGG 5820  
 DB 5761 TACAATGTTGAAGCTATGCAAAATAAAAACCATTTAATGATTTGGGTTTGAAGTAGG 5820  
 QY 5821 TAAAGTAAATGCTATGCAAGCAACCAAGCTGATTAATTTTGAAGAAATTAATCTTTG 5880  
 DB 5821 TAAAGTAAATGCTATGCAAGCAACCAAGCTGATTAATTTTGAAGAAATTAATCTTTG 5880  
 QY 5881 TCACCAATGACCAATGATATCTGCTATTAAGTTTCTAGCCAAATGATCAATGATAA 5940

DB 5881 TCACCAATGACCAATGATATCTGCTATTAAGTTTCTACGCAATGATCAATATAA 5940  
 QY 5941 AGTATTTCAATATCCATGATGGAAGCAATGATAATTAACCGGATTTGATGCTAAGG 6000  
 DB 5941 AGTATTTCAATATCCATGATGGAAGCAATGATAATTAACCGGATTTGATGCTAAGG 6000  
 QY 6001 TCTAAGTGTGCTAAATTCAAAGCAATTCGTTGAGGGTATGATAAAAAGTTAGACCTAGA 6060  
 DB 6001 TCTAAGTGTGCTAAATTCAAAGCAATTCGTTGAGGGTATGATAAAAAGTTAGACCTAGA 6060  
 QY 6061 AGATTAATCAAAAGCTATATCAAAAATGCAATTAACGCGAATAGTGATTTATCATTAAT 6120  
 DB 6061 AGATTAATCAAAAGCTATATCAAAAATGCAATTAACGCGAATAGTGATTTATCATTAAT 6120  
 QY 6121 ATTATGTAATGCTCAATCAGGGAGTGTGTTTGTGCACTAATTTTGTGCAAGAACTG 6180  
 DB 6121 ATTATGTAATGCTCAATCAGGGAGTGTGTTTGTGCACTAATTTTGTGCAAGAACTG 6180  
 QY 6181 ATCAATGAAAATGATGCTTATTTATGCAATGGAATTTTTCATACACATCAGGTGAT 6240  
 DB 6181 ATCAATGAAAATGATGCTTATTTATGCAATGGAATTTTTCATACACATCAGGTGAT 6240  
 QY 6241 GACAGAACTAACTATATTTGTGATATTAATTAAGGGTATGAAAATCCAAATCCAGAA 6300  
 DB 6241 GACAGAACTAACTATATTTGTGATATTAATTAAGGGTATGAAAATCCAAATCCAGAA 6300  
 QY 6301 TATCAACGAGATGGAAGGAGGAGTTTCAATATATCTTTGATATATPACTGTTCT 6360  
 DB 6301 TATCAACGAGATGGAAGGAGGAGTTTCAATATATCTTTGATATATPACTGTTCT 6360  
 QY 6361 AATTCACTATACCAACTGAGCTGATACAGCTCAATCCAGGTAAGAAAGTTATATT 6420  
 DB 6361 AATTCACTATACCAACTGAGCTGATACAGCTGATACAGCTCAATCCAGGTAAGAAAGTTATATT 6420  
 QY 6421 CCATCA 6426  
 DB 6421 CCATCA 6426  
 RESULT 2  
 AAAS7936  
 ID AAAS7936 standard; DNA; 392 Bp.  
 AC AAAS7936;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 DE  
 XX Candida albicans strain p36 Tca2 retrotransposon 5' region.  
 XX  
 KM Retrotransposon; pCa1; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KM 58S gene; group antigen; pol; protein; pol; aspartate protease; integrase;  
 KM reverse transcriptase; RNaseH; pseudonuc; readthrough translation;  
 KM scop codon suppression; gene delivery; gene therapy vector;  
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.  
 OS  
 XX Candida albicans; strain p36.  
 OS  
 PN W0200026397-A1.  
 PD 11-MAY-2000.  
 PD  
 PF 01-NOV-1999; 99WO-NZ000179.  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 PA (JANC) JANSSEN PHARM NV.  
 PI Luyten WHM, De Backer MD, Nelissen BW, Poulter RM;  
 XX MPI; 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an  
PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
of *Candida* in a sample.

PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
CC genome. In particular, the invention relates to the novel C. albicans  
CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
CC this retrotransposon, designated TCa2, and to the novel C. albicans  
CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
CC protease, integrase, reverse transcriptase (RT), and RNaseH. The gag and  
CC pol ORFs of pCal are in the same reading frame, separated only by a  
CC termination codon (TGA). Translation of the pol ORF occurs through the  
CC occasional readthrough suppression of the stop codon, which is mediated  
CC by the formation of a pseudoknot within the gag-pol mRNA. The  
CC retrotransposons of the invention can be used as vectors for in vitro or  
CC in vivo transformation and expression. They can thus be used for the  
CC delivery and expression of a therapeutic, immunological or immunogenic  
CC molecule (e.g., an antigen) and may also be used for eliciting an  
CC immunological response in a host organism. They are therefore useful in  
CC genetic vaccine compositions and for gene therapy, particularly where the  
CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
CC retrotransposons may be used to generate transgenic animals, to detect  
CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
CC assign functions to nucleotide sequences. Sequences AA57923-457936  
CC represents TCa2 retrotransposon 5' regions from a variety of C. albicans  
CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 392; DB 3; Length 392;  
Best Local Similarity 100.0%; Pred. No. 1.8e-69;  
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTGGTTGGCACTATTGTCGCAAACTGATCAAGAAAGTGGTATATGA 60  
DB 1 TGTGGTTGGCACTATTGTCGCAAACTGATCAAGAAAGTGGTATATGA 60  
QY 1 TGTGGTTGGCACTATTGTCGCAAACTGATCAAGAAAGTGGTATATGA 60  
DB 1 TGTGGTTGGCACTATTGTCGCAAACTGATCAAGAAAGTGGTATATGA 60  
QY 61 GAATGGAATTTTCCATCAACATCAGTGTACAGAACTAATATTTGTAGT 120  
DB 61 GAATGGAATTTTCCATCAACATCAGTGTACAGAACTAATATTTGTAGT 120  
QY 121 ATTAATTAAGGTATGAATTAACACATCCGCAATATTAAGCAATTAAGCGAGGAGT 180  
DB 121 ATTAATTAAGGTATGAATTAACACATCCGCAATATTAAGCAATTAAGCGAGGAGT 180  
QY 121 ATTAATTAAGGTATGAATTAACACATCCGCAATATTAAGCAATTAAGCGAGGAGT 180  
DB 121 ATTAATTAAGGTATGAATTAACACATCCGCAATATTAAGCAATTAAGCGAGGAGT 180  
QY 181 TCAATATATCTTGTGATATTAATCTGTTTCAATCACTATACCAACTAGACGTGT 240  
DB 181 TCAATATATCTTGTGATATTAATCTGTTTCAATCACTATACCAACTAGACGTGT 240  
QY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCAATTAAGTGAAGTGAATGAT 300  
DB 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCAATTAAGTGAAGTGAATGAT 300  
QY 301 AATCAATTCGTCGCAATTAAGGTATTAATTAATCAAGCCCAAGTTGATTTATGATT 360  
DB 301 AATCAATTCGTCGCAATTAAGGTATTAATTAATCAAGCCCAAGTTGATTTATGATT 360  
QY 361 GATAGTTTGAAGTTGAAGTGAAGAAATTTTC 392  
DB 361 GATAGTTTGAAGTTGAAGTGAAGAAATTTTC 392

RESULT 3  
AA57928  
ID AA57928 standard; DNA; 392 BP.

XX AA57928;  
XX 15-SEP-2003 (revised)  
DT 10-OCT-2000 (first entry)  
XX  
XX  
DE *Candida* albicans strain ATC-2 TCa2 retrotransposon 5' region.

XX Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;  
XX gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;  
XX reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
XX stop codon suppression; gene delivery; gene therapy vector;  
XX genetic vaccine composition; immunogenic; transgenic animal; de.

XX *Candida* albicans; strain ATC-2.

XX WO200026397-A1.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ000179.

XX 30-OCT-1998; 98CA-02249046.

XX 30-OCT-1998; 98US-0106342P.

XX (JANC) JANSSEN PHARM NV.

XX Luyten WHML, De Backer MD, Nelissen BM, Poulier RTM;  
XX WPI: 2000-365640/31.

XX Novel retrotransposon expression vectors useful for expressing an  
PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
of *Candida* in a sample.

XX Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
CC genome. In particular, the invention relates to the novel C. albicans  
CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
CC this retrotransposon, designated TCa2, and to the novel C. albicans  
CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
CC pol ORFs of pCal are in the same reading frame, separated only by a  
CC termination codon (TGA). Translation of the pol ORF occurs through the  
CC occasional readthrough suppression of the stop codon, which is mediated  
CC by the formation of a pseudoknot within the gag-pol mRNA. The  
CC retrotransposons of the invention can be used as vectors for in vitro or  
CC in vivo transformation and expression. They can thus be used for the  
CC delivery and expression of a therapeutic, immunological or immunogenic  
CC molecule (e.g., an antigen) and may also be used for eliciting an  
CC immunological response in a host organism. They are therefore useful in  
CC genetic vaccine compositions and for gene therapy, particularly where the  
CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
CC retrotransposons may be used to generate transgenic animals, to detect  
CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
CC assign functions to nucleotide sequences. Sequences AA57923-457936  
CC represents TCa2 retrotransposon 5' regions from a variety of C. albicans  
CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 388.8; DB 3; Length 392;  
Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGTTGGCACTATTGTCGCAAACTGATCAAGAAAGTGGTATATGA 60  
DB 1 TGTGGTTGGCACTATTGTCGCAAACTGATCAAGAAAGTGGTATATGA 60

Db 1 TGTGTGTTTGACACTATTTTGTGTCAGAACTGATCATGATAAATGATGTATTATGA 60  
 QY 61 GAATGGAAAAATTTTTCATCACAATCAGGTGATGACAGAACTAACTATATTGTGTA 120  
 Db 61 GAATGGAAAAATTTTTCATCACAATCAGGTGATGACAGAACTAACTATATTGTGTA 120  
 QY 121 ATAATTAAGGATGATGAAATACCAACATCCAGATATCAAGAGATAGAGAGAGATT 180  
 Db 121 ATAATTAAGGATGATGAAATACCAACATCCAGATATCAAGAGATAGAGAGAGATT 180  
 QY 181 TCAATATATATCTTGTGATATATACTCGTTCTTAATTCAGTATACACAACTAGACG 240  
 Db 181 TCAATATATATCTTGTGATATATACTCGTTCTTAATTCAGTATACACAACTAGACG 240  
 QY 241 ACAGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 Db 241 ACAGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 QY 301 AATCATTTTCGTCACAAATTAAGCGTTGATATAATTCAGTCTCAGATTTGATTA 360  
 Db 301 AATCATTTTCGTCACAAATTAAGCGTTGATATAATTCAGTCTCAGATTTGATTA 360  
 QY 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 Db 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 RESULT 4  
 ID AAA57931 standard; DNA; 392 BP.  
 XX AAA57931;  
 AC 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain Fl6-1 Tca2 retrotransposon 5' region.  
 XX  
 KM Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KM gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KM reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KM stop codon suppression; gene delivery; gene therapy vector;  
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS Candida albicans; strain Fl6-1.  
 XX  
 PN WO20026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JUNC) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;  
 DR WPI; 2000-365640/31.  
 XX  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pCal (AAA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans

CC hCG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AAA57923-457936  
 CC represent Tca2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;  
 Query Match 6.1%; Score 388.8; DB 3; Length 392;  
 Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 TGTGTGTTTGACACTATTTTGTGTCAGAACTGATCATGATAAATGATGTATTATGA 60  
 Db 1 TGTGTGTTTGACACTATTTTGTGTCAGAACTGATCATGATAAATGATGTATTATGA 60  
 QY 61 GAATGGAAAAATTTTTCATCACAATCAGGTGATGACAGAACTAACTATATTGTGTA 120  
 Db 61 GAATGGAAAAATTTTTCATCACAATCAGGTGATGACAGAACTAACTATATTGTGTA 120  
 QY 121 ATAATTAAGGATGATGAAATACCAACATCCAGATATCAAGAGATAGAGAGAGATT 180  
 Db 121 ATAATTAAGGATGATGAAATACCAACATCCAGATATCAAGAGATAGAGAGAGATT 180  
 QY 181 TCAATATATATCTTGTGATATATACTCGTTCTTAATTCAGTATACACAACTAGACG 240  
 Db 181 TCAATATATATCTTGTGATATATACTCGTTCTTAATTCAGTATACACAACTAGACG 240  
 QY 241 ACAGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 Db 241 ACAGCTCAATCTCAGTAAAGAAAGTTTATATTCATCAGATTAAGAGTGAATGAT 300  
 QY 301 AATCATTTTCGTCACAAATTAAGCGTTGATATAATTCAGTCTCAGATTTGATTA 360  
 Db 301 AATCATTTTCGTCACAAATTAAGCGTTGATATAATTCAGTCTCAGATTTGATTA 360  
 QY 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 Db 361 GATAGTTTCGAGATTGGAAGGTACAGAAATTTTC 392  
 RESULT 5  
 ID AAA57923 standard; DNA; 392 BP.  
 XX AAA57923;  
 AC 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain SGV-1 Tca2 retrotransposon 5' region.  
 XX  
 KM Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KM gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KM reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KM stop codon suppression; gene delivery; gene therapy vector;  
 KM genetic vaccine composition; immunogenic; transgenic animal; ds.

```

XX OS Candida albicans; strain SGY-1.
XX PN WO200026397-A1.
XX PD 11-MAY-2000.
XX PF 01-NOV-1999; 99WO-NZ000179.
XX PR 30-OCT-1998; 98CA-02249046.
XX PR 30-OCT-1998; 98US-0106342P.
XX PA (JUNC ) JANSSEN PHARM NV.
XX PI Luyten WHML, De Backer MD, Nelissen BOM, Poulter RTM;
XX DR WPI; 2000-365640/31.
XX PT Novel retrotransposon expression vectors useful for expressing an
XX PT antigen, epitope or therapeutic agent, or detecting genes or the presence
XX PT of Candida in a sample.
XX PS Example 9; Fig 9; 204pp; English.
XX CC The invention relates to novel retrotransposons from the yeast Candida
XX CC albicans which have a copy number of 40-150, preferably 50-100 copies per
XX CC genome. In particular, the invention relates to the novel C. albicans
XX CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of
XX CC this retrotransposon, designated TCa2, and to the novel C. albicans
XX CC retrotransposons 1-28. pCal was initially isolated from C. albicans
XX CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises
XX CC identical 280 bp long terminal repeats (LTRs) and two open reading frames
XX CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the
XX CC second ORF encodes a polypeptide (pol) consisting of an aspartate
XX CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and
XX CC pol ORFs of pCal are in the same reading frame, separated only by a
XX CC termination codon (TGA). Translation of the pol ORF occurs through the
XX CC occasional readthrough suppression of the stop codon, which is mediated
XX CC by the formation of a pseudoknot within the gag-pol mRNA. The
XX CC retrotransposons of the invention can be used as vectors for in vitro or
XX CC in vivo transformation and expression. They can thus be used for the
XX CC delivery and expression of a therapeutic, immunological or immunogenic
XX CC molecule (e.g., an antigen) and may also be used for eliciting an
XX CC immunological response in a host organism. They are therefore useful in
XX CC genetic vaccine compositions and for gene therapy, particularly where the
XX CC use of retroviral vectors is unsafe or undesirable. Additionally, the
XX CC retrotransposons may be used to generate transgenic animals, to detect
XX CC the presence of Candida in a sample, to detect and disrupt genes, and to
XX CC assign functions to nucleotide sequences. Sequences AA57923-AA57936
XX CC represents TCa2 retrotransposon 5' regions from a variety of C. albicans
XX CC strains. (Updated on 15-SEP-2003 to standardise OS field)
XX SQ Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;
XX
XX Query Match 6.1%; Score 388.8; DB 3; Length 392;
XX Best Local Similarity 99.5%; Pred. No. 7.9e-69;
XX Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 1 TGTGGTTGTCATATTTGTGTGCAAGAACTGATCAATGAAATGATGTTATTATGA 60
XX DB 1 TGTGGTTGTCATATTTGTGTGCAAGAACTGATCAATGAAATGATGTTATTATGA 60
XX
XX 61 GAATGAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATTGTTGAGT 120
XX DB 61 GAATGAAAATTTTCCATCACAATCAGTGATGACAGAACTAACTATTGTTGAGT 120
XX
XX 121 ATAATTAAGGATGAATATCAACATCCCAATATCAAGATAGAGAGAGAGATT 180
XX DB 121 ATAATTAAGGATGAATATCAACATCCCAATATCAAGATAGAGAGAGAGATT 180
XX
XX 181 TCAATATATATCTTGGAATATATTAATCTGTTCTTAATTCATATACACTTACAGCTGT 240
XX DB 181 TCAATATATATCTTGGAATATATTAATCTGTTCTTAATTCATATACACTTACAGCTGT 240

```

```

XX QY 241 ACAGCTATATCTCAGGTAAAGAAAGTTTATTCATCCATGATTAGAGTGCATAGTAT 300
XX DB 241 ACAGCTATATCTCAGGTAAAGAAAGTTTATTCATCCATGATTAGAGTGCATAGTAT 300
XX
XX QY 301 AATCATTTGTCGCCAAATATAGCTTGTATTAATTCAGTCCATGATTGATATTCATT 360
XX DB 301 AATCATTTGTCGCCAAATATAGCTTGTATTAATTCAGTCCATGATTGATATTCATT 360
XX
XX QY 361 GATAGTTTCGAAGTTTGAGAGGTACAGAAATTC 392
XX DB 361 GATAGTTTCGAAGTTTGAGAGGTACAGAAATTC 392
XX
XX RESULT 6
XX ID AA57932 standard; DNA; 392 BP.
XX AC AA57932;
XX XX
XX DT 15-SEP-2003 (revised)
XX DT 10-OCT-2000 (first entry)
XX XX
XX DE Candida albicans strain Fl6-2 TCa2 retrotransposon 5' region.
XX XX
XX XX Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;
XX XX gag gene; group antigen; polypeptide; pol; aspartate protease; integrase;
XX XX reverse transcriptase; RNaseH; pseudoknot; readthrough translation;
XX XX stop codon suppression; gene delivery; gene therapy vector;
XX XX genetic vaccine composition; immunogenic; transgenic animal; de.
XX XX
XX OS Candida albicans; strain Fl6-2.
XX PN WO200026397-A1.
XX PD 11-MAY-2000.
XX PF 01-NOV-1999; 99WO-NZ000179.
XX PR 30-OCT-1998; 98CA-02249046.
XX PR 30-OCT-1998; 98US-0106342P.
XX PA (JUNC ) JANSSEN PHARM NV.
XX PI Luyten WHML, De Backer MD, Nelissen BOM, Poulter RTM;
XX DR WPI; 2000-365640/31.
XX PT Novel retrotransposon expression vectors useful for expressing an
XX PT antigen, epitope or therapeutic agent, or detecting genes or the presence
XX PT of Candida in a sample.
XX PS Example 9; Fig 9; 204pp; English.
XX CC The invention relates to novel retrotransposons from the yeast Candida
XX CC albicans which have a copy number of 40-150, preferably 50-100 copies per
XX CC genome. In particular, the invention relates to the novel C. albicans
XX CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of
XX CC this retrotransposon, designated TCa2, and to the novel C. albicans
XX CC retrotransposons 1-28. pCal was initially isolated from C. albicans
XX CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises
XX CC identical 280 bp long terminal repeats (LTRs) and two open reading frames
XX CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the
XX CC second ORF encodes a polypeptide (pol) consisting of an aspartate
XX CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and
XX CC pol ORFs of pCal are in the same reading frame, separated only by a
XX CC termination codon (TGA). Translation of the pol ORF occurs through the
XX CC occasional readthrough suppression of the stop codon, which is mediated
XX CC by the formation of a pseudoknot within the gag-pol mRNA. The
XX CC retrotransposons of the invention can be used as vectors for in vitro or
XX CC in vivo transformation and expression. They can thus be used for the
XX CC delivery and expression of a therapeutic, immunological or immunogenic
XX CC molecule (e.g., an antigen) and may also be used for eliciting an

```

CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-AS7936  
 CC represents Tca2 retrotransposon 5' regions from a variety of *C. albicans*  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 388.8; DB 3; Length 392;  
 Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGTTGGTCACTATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 DB 1 TGTGGTTGGTCACTATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 QY 61 GAATGAAAAATTTTCCATCACAATCAGGTGATGACAGAACTAACTATATGTAGT 120  
 DB 61 GAATGAAAAATTTTCCATCACAATCAGGTGATGACAGAACTAACTATATGTAGT 120  
 QY 121 ATAAATTAAGGGTATTAATTAATCAATCCCAAGATTAAGAGATTAAGAGGAGT 180  
 DB 121 ATAAATTAAGGGTATTAATTAATCAATCCCAAGATTAAGAGATTAAGAGGAGT 180  
 QY 181 TCAATATATATCTTGATATATAAATCTGTTCTAATTCATATACCAATAGACGTGT 240  
 DB 181 TCAATATATATCTTGATATATAAATCTGTTCTAATTCATATACCAATAGACGTGT 240  
 QY 241 ACACGCTCAATTCAGGTAAAGAAAGTTTATTTCCATCAATTAAGAGTATGAT 300  
 DB 241 ACACGCTCAATTCAGGTAAAGAAAGTTTATTTCCATCAATTAAGAGTATGAT 300  
 QY 301 AATCATTTGCTCCCAATTAAGGTGTATTAATTCAGTCTCAGATTGTATTTAGT 360  
 DB 301 AATCATTTGCTCCCAATTAAGGTGTATTAATTCAGTCTCAGATTGTATTTAGT 360  
 QY 361 GATAGTTTTCGAGTTTGAAGGTACAGAAATTTTC 392  
 DB 361 GATAGTTTTCGAGTTTGAAGGTACAGAAATTTTC 392

RESULT 7

AA57927  
 ID AA57927 standard; DNA; 392 BP.

AC AA57927;  
 XX  
 DT 15-SEP-2003 (Revised)  
 DT 10-OCT-2000 (first entry)

DE *Candida albicans* strain ATC-1 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.

XX *Candida albicans*; strain ATC-1.

XX MO200026397-AI.

XX 11-MAY-2000.

XX 01-NOV-1999; 99WO-NZ000179.

XX 30-OCT-1998; 98CA-02249046.

XX 30-OCT-1998; 98US-0106342P.

XX (JANSEN PHARM NV.

XX Luyten WML, De Backer MD, Nelissen BJM, Poulsen RTM;  
 PI WPI; 2000-365640/31.  
 XX  
 DR Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of *Candida* in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.

XX The invention relates to novel retrotransposons from the yeast *Candida*  
 CC *albicans* which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel *C. albicans*  
 CC Ty1/copia retrotransposon pCal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel *C. albicans*  
 CC retrotransposons 1-28. pCal was initially isolated from *C. albicans*  
 CC H0G1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-AS7936  
 CC represents Tca2 retrotransposon 5' regions from a variety of *C. albicans*  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)

XX Sequence 392 BP; 136 A; 56 C; 72 G; 128 T; 0 U; 0 Other;

Query Match 6.1%; Score 388.8; DB 3; Length 392;  
 Best Local Similarity 99.5%; Pred. No. 7.9e-69;  
 Matches 390; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTGGTTGGTCACTATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 DB 1 TGTGGTTGGTCACTATTTTGTGCAGAACTGATCAATGAAATGATGTTTATGA 60  
 QY 61 GAATGAAAAATTTTCCATCACAATCAGGTGATGACAGAACTAACTATATGTAGT 120  
 DB 61 GAATGAAAAATTTTCCATCACAATCAGGTGATGACAGAACTAACTATATGTAGT 120  
 QY 121 ATAAATTAAGGGTATTAATTAATCAATCCCAAGATTAAGAGATTAAGAGGAGT 180  
 DB 121 ATAAATTAAGGGTATTAATTAATCAATCCCAAGATTAAGAGATTAAGAGGAGT 180  
 QY 181 TCAATATATATCTTGATATATAAATCTGTTCTAATTCATATACCAATAGACGTGT 240  
 DB 181 TCAATATATATCTTGATATATAAATCTGTTCTAATTCATATACCAATAGACGTGT 240  
 QY 241 ACACGCTCAATTCAGGTAAAGAAAGTTTATTTCCATCAATTAAGAGTATGAT 300  
 DB 241 ACACGCTCAATTCAGGTAAAGAAAGTTTATTTCCATCAATTAAGAGTATGAT 300  
 QY 301 AATCATTTGCTCCCAATTAAGGTGTATTAATTCAGTCTCAGATTGTATTTAGT 360  
 DB 301 AATCATTTGCTCCCAATTAAGGTGTATTAATTCAGTCTCAGATTGTATTTAGT 360  
 QY 361 GATAGTTTTCGAGTTTGAAGGTACAGAAATTTTC 392  
 DB 361 GATAGTTTTCGAGTTTGAAGGTACAGAAATTTTC 392

## RESULT 8

AAA57925 standard; DNA, 392 BP.  
 ID AAA57925 standard; DNA, 392 BP.  
 AC AAA57925;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain SCS-1 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS Candida albicans; strain SCS-1.  
 XX  
 PN WO200026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JANC) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHM, De Backer MD, Nelissen BJM, Poulter RTM;  
 PI WPI; 2000-365640/31.  
 XX  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pCal (AAA57920) and to the integrated form of  
 CC this retrotransposon, designated Tca2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
 CC hOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AAA57923-A57936  
 CC represent Tca2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 XX  
 SQ Sequence 392 BP; 136 A; 56 C; 71 G; 129 T; 0 U; 0 Other;

Query Match 6 0%; Score 387.2; DB 3; Length 392;  
 Best Local Similarity 99.2%; Pred. No. 1.7e-68;

Matches 389; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGTGGTTTGGCACTATTTTGTGTCAGAACTGATCAATGAATGATGTTATATGA 60  
 DB 1 TGTGGTTTGGCACTATTTTGTGTCAGAACTGATCAATGAATGATGTTATATGA 60  
 QY 61 GATGGAAAAATTTTTCATCATCATGAGTGATGACAGAACTAATATATGTTAGT 120  
 DB 61 GATGGAAAAATTTTTCATCATCATGAGTGATGACAGAACTAATATATGTTAGT 120  
 QY 121 ATTAATAAGGATGATAATACCAATCCAGATATCAACGATAGAAAGGAGAGT 180  
 DB 121 ATTAATAAGGATGATAATACCAATCCAGATATCAACGATAGAAAGGAGAGT 180  
 QY 181 TCAATATATATCTTGTGATATATATCTTGTGATATATCACTAATCACTAGAGT 240  
 DB 181 TCAATATATATCTTGTGATATATATCTTGTGATATATCACTAATCACTAGAGT 240  
 QY 241 ACAGGCTCAATCTCAGGTAAGTAAGTTATATCCATCATGATAGAGTCAATGAT 300  
 DB 241 ACAGGCTCAATCTCAGGTAAGTAAGTTATATCCATCATGATAGAGTCAATGAT 300  
 QY 301 AATCATTTTCGTCCTCAATTTAGCGTTGTATATATTCAGTCTCAGATTTTATAT 360  
 DB 301 AATCATTTTCGTCCTCAATTTAGCGTTGTATATATTCAGTCTCAGATTTTATAT 360  
 QY 361 GATAGTTTCGAGATTTGAAGGTACAGAAATTTTC 392  
 DB 361 GATAGTTTCGAGATTTGAAGGTACAGAAATTTTC 392

## RESULT 9

AAA57933 standard; DNA, 392 BP.

AAA57933;  
 ID AAA57933 standard; DNA, 392 BP.  
 AC AAA57933;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain 759-1 Tca2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCal; Tca2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS Candida albicans; strain 759-1.  
 XX  
 PN WO200026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JANC) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHM, De Backer MD, Nelissen BJM, Poulter RTM;  
 PI WPI; 2000-365640/31.  
 XX  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per









PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX (JANC ) JANSSEN PHARM NV.  
 PA  
 XX Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;  
 XX MPI: 2000-365640/31.  
 DR  
 XX  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 PS  
 XX Example 9; Fig 9; 204pp; English.  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pcal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated TCa2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
 CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pcal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-A57936  
 CC represents TCa2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 XX  
 SS Sequence 392 BP; 138 A; 59 C; 71 G; 124 T; 0 U; 0 Other;  
 Query Match 6.0%; Score 384; DB 3; Length 392;  
 Best Local Similarity 98.7%; Pred. No. 7.4e-68;  
 Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Oy 1 TGTGGTTTGGCACTATTTTGTGACAAATGATCATATAAATGATGTTTATGA 60  
 Db 1 TGTGGTTTGGCACTATTTTGTGACAAATGATCATATAAATGATGTTTATGA 60  
 Oy 61 GAATGAATAATTTTCCATCACATCAGTGATGACAGAACTAACTATTGTGTAGT 120  
 Db 61 GAATGAATAATTTTCCATCACATCAGTGATGACAGAACTAACTATTGTGTAGT 120  
 Oy 121 ATTAATTAAGGTATGAATACCAACATCCAGAAATATCAACGAGATGAGAGAGATT 180  
 Db 121 ATTAATTAAGGTATGAATACCAACATCCAGAAATATCAACGAGATGAGAGAGATT 180  
 Oy 121 ATTAATTAAGGTATGAATACCAACATCCAGAAATATCAACGAGATGAGAGAGATT 180  
 Db 121 ATTAATTAAGGTATGAATACCAACATCCAGAAATATCAACGAGATGAGAGAGATT 180  
 Oy 181 TCAATATATATCTTGTGAAATATTAATCTCGTTTCAATCACTATACCACTACAGTGT 240  
 Db 181 TCAATATATATCTTGTGAAATATTAATCTCGTTTCAATCACTATACCACTACAGTGT 240  
 Oy 241 ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTGAGAGTGTATGAT 300  
 Db 241 ACACGCTCAATCTCAGTAAAGAAAGTTATATTCATCAGATTGAGAGTGTATGAT 300  
 Oy 301 AATCATTTTCGCCCAATTAAGCGTTGTATAAATCAGTCCCAATTTGATTTATGAT 360  
 Db 301 AATCATTTTCGCCCAATTAAGCGTTGTATAAATCAGTCCCAATTTGATTTATGAT 360

Oy 361 GATAGTTTGGAAAGTTTGAAGCTACAGAAATTTC 392  
 Db 361 GATAGTTTGGAAAGTTTGAAGCTACAGAAATTTC 392  
 RESULT 13  
 ID AA57935 standard; DNA; 392 BP.  
 XX  
 AC AA57935;  
 XX  
 DT 15-SEP-2003 (revised)  
 DT 10-OCT-2000 (first entry)  
 XX  
 DE Candida albicans strain p30 TCa2 retrotransposon 5' region.  
 XX  
 KW Retrotransposon; pCal; TCa2; Ty1; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; ds.  
 XX  
 OS Candida albicans; strain p30.  
 XX  
 PN WO200026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WC-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JANC ) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;  
 XX  
 DR MPI: 2000-365640/31.  
 XX  
 PT Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9; Fig 9; 204pp; English.  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Ty1/copia retrotransposon pcal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated TCa2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
 CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polypeptide (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pcal are in the same reading frame, separated only by a  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-A57936  
 CC represents TCa2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 XX

SQ Sequence 392 BP; 139 A; 55 C; 71 G; 127 T; 0 U; 0 Other;

Query Match 6.0%; Score 384; DB 3; Length 392;  
Best Local Similarity 98.7%; Pred. No. 7,4e-68;  
Matches 387; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
QY 1 TGTGGTTTGGCACTATTTTGTGTGCAAACTGATCAATGAAATGATGTTATTTATGA 60
DB 1 TGTGGTTTGGCACTATTTTGTGTGCAAACTGATCAATGAAATGATGTTATTTATGA 60
QY 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
QY 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 ATAAATAAGGGTATGAAATACCAACATCCAGAAATATCAACGATGAAAGGAGATTT 180
DB 121 ATAAATAAGGGTATGAAATACCAACATCCAGAAATATCAACGATGAAAGGAGATTT 180
QY 121 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
DB 121 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
QY 181 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
DB 181 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
QY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATGATGATGATGAT 300
DB 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATGATGATGATGAT 300
QY 301 AATCATTTTCCGCCAATTTAGCGTTGTATTAATTCAGTCTCTGATTTGATTTGATTT 360
DB 301 AATCATTTTCCGCCAATTTAGCGTTGTATTAATTCAGTCTCTGATTTGATTTGATTT 360
QY 301 AATCATTTTCCGCCAATTTAGCGTTGTATTAATTCAGTCTCTGATTTGATTTGATTT 360
DB 301 AATCATTTTCCGCCAATTTAGCGTTGTATTAATTCAGTCTCTGATTTGATTTGATTT 360
QY 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTTTC 392
DB 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTTTC 392
```

RESULT 14  
ID AA57924  
AA57924 standard; DNA; 392 BP.

AC AA57924;

DT 15-SEP-2003 (revised)  
DT 10-OCT-2000 (first entry)

DE Candida albicans strain SGY-2 TCa2 retrotransposon 5' region.

KM Retrotransposon; pCal; TCa2; Tyl; copias; long terminal repeat; LTR;  
KX gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
KM reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
KM stop codon suppression; gene delivery; gene therapy vector;  
KM genetic vaccine composition; immunogenic; transgenic animal; ds.

OS Candida albicans; strain SGY-2.

FN WO2000026397-A1.

PD 11-MAY-2000.

PF 01-NOV-1999; 99WO-NZ000179.

PR 30-OCT-1998; 98CA-02249046.

PR 30-OCT-1998; 98US-0106342P.

PA (JANC) JANSSEN PHARM NV.

PI Luyten WHML, De Backer MD, Nelissen BJM, Poultier RTM;

DE WPI; 2000-365640/31.

PT Novel retrotransposon expression vectors useful for expressing an  
PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
PT of Candida in a sample.

PS Example 9; Fig 9; 204pp; English.

CC The invention relates to novel retrotransposons from the yeast *Candida*  
CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
CC genome. In particular, the invention relates to the novel *C. albicans*  
CC Tyl/copia retrotransposon pCal (AA57920), and to the integrated form of  
CC this retrotransposon, designated TCa2, and to the novel *C. albicans*  
CC retrotransposons 1-28. pCal was initially isolated from *C. albicans*  
CC HOG1042 and has a copy number of 50-100 copies per cell. It comprises  
CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
CC pol ORFs of pCal are in the same reading frame, separated only by a  
CC termination codon (TGA). Translation of the pol ORF occurs through the  
CC occasional readthrough suppression of the stop codon, which is mediated  
CC by the formation of a pseudoknot within the gag-pol mRNA. The in vitro or  
CC retrotransposons of the invention can be used as vectors for the  
CC in vivo transformation and expression. They can thus be used for the  
CC delivery and expression of a therapeutic, immunological or immunogenic  
CC molecule (e.g., an antigen) and may also be used for eliciting an  
CC immunological response in a host organism. They are therefore useful in  
CC genetic vaccine compositions and for gene therapy, particularly where the  
CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
CC retrotransposons may be used to generate transgenic animals, to detect  
CC the presence of *Candida* in a sample, to detect and disrupt genes, and to  
CC assign functions to nucleotide sequences. Sequences AA57923-AA7936  
CC represent TCa2 retrotransposon 5' regions from a variety of *C. albicans*  
CC strains. (Updated on 15-SEP-2003 to standardise Os field)

SQ Sequence 392 BP; 135 A; 57 C; 74 G; 126 T; 0 U; 0 Other;

Query Match 6.0%; Score 382.4; DB 3; Length 392;  
Best Local Similarity 98.5%; Pred. No. 1.1e-67;  
Matches 386; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
QY 1 TGTGGTTTGGCACTATTTTGTGTGCAAACTGATCAATGAAATGATGTTATTTATGA 60
DB 1 TGTGGTTTGGCACTATTTTGTGTGCAAACTGATCAATGAAATGATGTTATTTATGA 60
QY 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
QY 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
DB 61 GAATGAAAAATTTTCCATCATCATCATGATGATGATGATGATGATGATGATGATGAT 120
QY 121 ATAAATAAGGGTATGAAATACCAACATCCAGAAATATCAACGATGAAAGGAGATTT 180
DB 121 ATAAATAAGGGTATGAAATACCAACATCCAGAAATATCAACGATGAAAGGAGATTT 180
QY 121 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
DB 121 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
QY 181 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
DB 181 TCAATATATATCTGTGTAATATATATCTGTGTCAATCTCTATACCACTGACGTGT 240
QY 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATGATGATGATGAT 300
DB 241 ACACGCTCAATCTCAGGTAAAGAAAGTTATATTCATCATCATGATGATGATGATGAT 300
QY 301 AATCATTTTCCGCCAATTTAGCGTTGTATTAATTCAGTCTCTGATTTGATTTGATTT 360
DB 301 AATCATTTTCCGCCAATTTAGCGTTGTATTAATTCAGTCTCTGATTTGATTTGATTT 360
QY 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTTTC 392
DB 361 GATAGTTTCGAAGTTTGAAGTACAGAAATTTTC 392
```

RESULT 15  
ID AA57926  
AA57926 standard; DNA; 392 BP.

AC AA57926;

DT 15-SEP-2003 (revised)  
DT 10-OCT-2000 (first entry)

XX DE Candida albicans strain SC5-2 Tca2 retrotransposon 5' region.  
 XX XX Retrotransposon; pCal; Tca2; Tyl; copia; long terminal repeat; LTR;  
 KW gag gene; group antigen; polyprotein; pol; aspartate protease; integrase;  
 KW reverse transcriptase; RNaseH; pseudoknot; readthrough translation;  
 KW stop codon suppression; gene delivery; gene therapy vector;  
 KW genetic vaccine composition; immunogenic; transgenic animal; de.  
 XX  
 OS Candida albicans; strain SC5-2.  
 XX  
 PN WO20026397-A1.  
 XX  
 PD 11-MAY-2000.  
 XX  
 PF 01-NOV-1999; 99WO-NZ000179.  
 XX  
 PR 30-OCT-1998; 98CA-02249046.  
 PR 30-OCT-1998; 98US-0106342P.  
 XX  
 PA (JANCO) JANSSEN PHARM NV.  
 XX  
 PI Luyten WHM, De Backer MD, Nelissen BJM, Poulter RTM;  
 XX  
 DR WPI; 2000-365640/31.  
 XX  
 XX Novel retrotransposon expression vectors useful for expressing an  
 PT antigen, epitope or therapeutic agent, or detecting genes or the presence  
 PT of Candida in a sample.  
 XX  
 PS Example 9, Fig 9; 204pp; English.  
 XX  
 CC The invention relates to novel retrotransposons from the yeast Candida  
 CC albicans which have a copy number of 40-150, preferably 50-100 copies per  
 CC genome. In particular, the invention relates to the novel C. albicans  
 CC Tyl/copia retrotransposon pCal (AA57920), and to the integrated form of  
 CC this retrotransposon, designated TCA2, and to the novel C. albicans  
 CC retrotransposons 1-28. pCal was initially isolated from C. albicans  
 CC hOG1042 and has a copy number of 50-100 copies per cell. It comprises  
 CC identical 280 bp long terminal repeats (LTRs) and two open reading frames  
 CC (ORFs). The first ORF encodes a gag (group antigen) protein, and the  
 CC second ORF encodes a polyprotein (pol) consisting of an aspartate  
 CC protease, integrase, reverse transcriptase (RT) and RNaseH. The gag and  
 CC pol ORFs of pCal are in the same reading frame, separated only by a  
 CC termination codon (TGA). Translation of the pol ORF occurs through the  
 CC occasional readthrough suppression of the stop codon, which is mediated  
 CC by the formation of a pseudoknot within the gag-pol mRNA. The  
 CC retrotransposons of the invention can be used as vectors for in vitro or  
 CC in vivo transformation and expression. They can thus be used for the  
 CC delivery and expression of a therapeutic, immunological or immunogenic  
 CC molecule (e.g., an antigen) and may also be used for eliciting an  
 CC immunological response in a host organism. They are therefore useful in  
 CC genetic vaccine compositions and for gene therapy, particularly where the  
 CC use of retroviral vectors is unsafe or undesirable. Additionally, the  
 CC retrotransposons may be used to generate transgenic animals, to detect  
 CC the presence of Candida in a sample, to detect and disrupt genes, and to  
 CC assign functions to nucleotide sequences. Sequences AA57923-457936  
 CC represents Tca2 retrotransposon 5' regions from a variety of C. albicans  
 CC strains. (Updated on 15-SEP-2003 to standardise OS field)  
 CC  
 SQ Sequence 392 BP; 137 A; 57 C; 71 G; 127 T; 0 U; 0 Other;

Query Match 5.9%; Score 380.8; DB 3; Length 392;

Best Local Similarity 98.2%; Pred. No. 3.3e-67; Matches 385; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 TGTGGTTGTCACATATTTGTGTCAGAACTGATCAGTGAATGATGGTTATATGA 60  
 Db 1 TGTGGTTGTCACATATTTGTGTCAGAACTGATCAGTGAATGATGGTTATATGA 60  
 QY 61 GAATGAAAATTTTCCATCACAATCGAGTGAATGACAGAACTAACTATTTGTAGT 120  
 Db 61 GAATGAAAATTTTCCATCACAATCGAGTGAATGACAGAACTAACTATTTGTAGT 120

QY 121 ATAAATAGGTAATGAATACCAATCCAGATATCAAGAGATAGAGGAGAGT 180  
 Db 121 ATAAATAGGTAATGAATACCAATCCAGATATCAAGAGATAGAGGAGAGT 180  
 QY 181 TCAATATATCTTGTGAATTAATACCTGTTCTAATTCATACACAACTAGACGTGT 240  
 Db 181 TCAATATATCTTGTGAATTAATACCTGTTCTAATTCATACACAACTAGACGTGT 240  
 QY 241 ACAAGCTCATCTCAGGTAAAGAACTTAATTCATCAGATTGAAGTCATAGTGT 300  
 Db 241 ACAAGCTCATCTCAGGTAAAGAACTTAATTCATCAGATTGAAGTCATAGTGT 300  
 QY 301 AATCATTTGTCGCCAATTAGCGTTGTATTAATTCAGTCCCTCAGATTGTATTTGATT 360  
 Db 301 AATCATTTGTCGCCAATTAGCGTTGTATTAATTCAGTCCCTCAGATTGTATTTGATT 360  
 QY 361 GATAGTTTGAAGTTTGAAGGTACAGAAATTTT 392  
 Db 361 GATAGTTTGAAGTTTGAAGGTACAGAAATTTT 392

Search completed: April 11, 2004, 14:02:32  
 Job time : 1504 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 13:03:47 ; Search time 289 Seconds

(without alignments)  
12339,496 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426

Sequence: 1 tggctgttgcacatattc.....agaagttatattccatca 6426

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	65.6	1.0	7218	1 US-08-232-463-14	Sequence 14, Appl
2	58.2	0.9	640681	4 US-09-790-988-1	Sequence 1, Appl
3	58	0.9	240	1 US-08-628-417-6	Sequence 6, Appl
4	57.2	0.9	2394	4 US-09-800-729-33	Sequence 33, Appl
C 5	56.8	0.9	5219	4 US-10-204-708-52	Sequence 52, Appl
6	56.4	0.9	2447	2 US-09-014-969-14	Sequence 14, Appl
7	56	0.9	5361	3 US-08-973-462-2	Sequence 2, Appl
8	56	0.9	6152	3 US-08-973-462-1	Sequence 1, Appl
C 9	55.6	0.9	396	4 US-09-640-173-53	Sequence 53, Appl
10	55.6	0.9	396	4 US-09-713-550-53	Sequence 53, Appl
11	55.6	0.9	11485	4 US-09-410-464-9	Sequence 9, Appl
12	55.4	0.9	2826	4 US-09-254-776B-50	Sequence 50, Appl
13	55.2	0.9	1696	4 US-09-835-811-1	Sequence 1, Appl
14	54.6	0.8	6660	2 US-08-771-602D-1	Sequence 1, Appl
15	54.6	0.8	6660	3 US-09-233-446B-1	Sequence 1, Appl
16	54.4	0.8	3275	4 US-09-370-838-151	Sequence 151, App
17	54	0.8	674	4 US-09-620-405B-465	Sequence 465, App
18	54	0.8	674	4 US-09-433-826B-465	Sequence 465, App
19	54	0.8	674	4 US-09-604-287A-465	Sequence 465, App
20	54	0.8	674	4 US-09-834-759-465	Sequence 465, App
C 21	53.6	0.8	5152	4 US-10-204-708-47	Sequence 47, Appl
22	53.2	0.8	1798	4 US-09-797-906-1	Sequence 1, Appl
23	53.2	0.8	8820	2 US-08-446-855A-1	Sequence 1, Appl
24	53.2	0.8	8920	3 US-09-150-741-1	Sequence 1, Appl
25	53	0.8	830	4 US-08-956-171E-715	Sequence 715, App
26	52.8	0.8	1784	4 US-09-601-198-21	Sequence 21, Appl
27	52.6	0.8	371	4 US-09-621-976-16048	Sequence 16048, A

C 28	52.4	0.8	5340	4 US-09-627-122-21	Sequence 21, Appl
29	52.2	0.8	14066	4 US-09-601-198-56	Sequence 56, Appl
30	52	0.8	2223	1 US-08-257-073-4	Sequence 4, Appl
31	52	0.8	30549	4 US-09-134-001C-322	Sequence 322, App
C 32	51.8	0.8	1020	4 US-09-328-475C-43	Sequence 43, Appl
33	51.4	0.8	1447	4 US-09-443-041A-27	Sequence 27, Appl
C 34	51	0.8	8607	4 US-10-204-708-72	Sequence 72, Appl
35	50.8	0.8	1866	4 US-09-601-198-153	Sequence 153, App
36	50.8	0.8	3095	6 US-08-280-443-1	Sequence 1, Appl
37	50.2	0.8	6671	1 US-08-457-459-1	Sequence 1, Appl
38	50.2	0.8	6671	1 US-08-555-678-1	Sequence 1, Appl
39	50.2	0.8	6671	5 PCT-US95-02275-1	Sequence 29, Appl
C 40	50.2	0.8	260	2 US-08-520-678A-29	Sequence 29, Appl
C 41	50	0.8	260	3 US-08-897-125-29	Sequence 10, Appl
C 42	50	0.8	260	3 US-09-601-537-10	Sequence 509, App
43	50	0.8	1481	4 US-09-134-001C-509	Sequence 9, Appl
44	50	0.8	1281	4 US-09-601-537-9	
45	50	0.8	4121	4 US-09-601-537-9	

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
City: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: BP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET INFORMATION:  
TELECOMMUNICATION INFORMATION: 30472/114 INMU  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZ9P-Fls  
US-08-232-463-14  
Query Match 1.0%; Score 65.6; DB 1; Length 7218;

1345 TCGGAGACCTGACACCATCTCCAAAAGAGAACCTTGAAATCAAGAAAGAGAAATTAAGA 1304

Page 3

RESULT 5  
US-10-204-708-52/c  
Sequence 52, Application US/10204708  
Patent No. 6677731  
GENERAL INFORMATION:  
APPLICANT: OLEK, Alexander  
APPLICANT: PIEPENROCK, Christian  
APPLICANT: BERLIN, Kurt  
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication  
TITLE OF INVENTION: By Assessing DNA Methylation  
FILE REFERENCE: 5013.1012  
CURRENT APPLICATION NUMBER: US/10/204,708  
CURRENT FILING DATE: 2003-05-06  
PRIOR APPLICATION NUMBER: PCP/EP01/03971  
PRIOR FILING DATE: 2001-04-06

1 Patent No. 5965397  
 2  
 3 GENERAL INFORMATION:  
 4 APPLICANT: Jacobs, Kenneth  
 5 APPLICANT: McCoy, John M.  
 6 APPLICANT: Lavallee, Edward R.  
 7 APPLICANT: Racie, Lisa A.  
 8 APPLICANT: Merberg, David  
 9 APPLICANT: Treacy, Maurice  
 10 APPLICANT: Spaulding, Vukki  
 11 APPLICANT: Agostino, Michael J.  
 12  
 13 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
 14  
 15 TITLE OF INVENTION: ENCODING THEM  
 16  
 17 NUMBER OF SEQUENCES: 32  
 18  
 19 CORRESPONDENCE ADDRESS:  
 20  
 21 ADDRESSEE: Genetics Institute, Inc.  
 22 STREET: 87 Cambridgepark Drive  
 23 CITY: Cambridge  
 24 STATE: MA  
 25  
 26 COUNTRY: U.S.A.  
 27  
 28 ZIP: 02140  
 29  
 30 COMPUTER READABLE FORM:  
 31  
 32 MEDIUM TYPE: Floppy disk  
 33  
 34 COMPUTER: IBM PC compatible  
 35  
 36 OPERATING SYSTEM: PC-DOS/MS-DOS  
 37  
 38 SOFTWARE: Patentin Release #1.0, Version #1.30  
 39  
 40 CURRENT APPLICATION DATA:  
 41  
 42 APPLICATION NUMBER: US/09/014,969  
 43  
 44 FILING DATE:  
 45

## CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:  
NAME: Springer, Suzanne A.  
REGISTRATION NUMBER: 41,323  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8284  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2447 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-09-014-969-14

## Query Match

Best Local Similarity 53.8%; Score 56.4; DB 2; Length 2447;  
Pred. No. 0.0018; Mismatches 97; Indels 0; Gaps 0;

Matches 114; Conservative 1; Mismatches 97; Indels 0; Gaps 0;

Db 1167 TTGACATTCAGACGACAAATATGACTTAATCTTATTTAGTTTACGACACGACAGAG 1226  
2203 TTGATTTTGTACTTTAATGACAAATTAACCTTTGGGAGAAAAA 2262  
Qy 1227 AACCAAAAGGAAACCGAGAGACTCCTGACACATCTCAAAAAGAACTGAAT 1286  
Db 2263 AA 2322  
Qy 1287 CAAGAAAGAAATAGAAACATCCAAATCAGATAACGATGAAGAAAGAAAG 1346  
Db 2323 AA 2382  
Qy 1347 AAAAAAAAAAACTCTGATGAAACAA 1378  
Db 2383 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2414

## RESULT 7

US-08-973-462-2  
Sequence 2, Application US/08973462B

Patent No. 6191270  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
APPLICANT: DAUBERSIES, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
CURRENT FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
EARLIER FILING DATE: 1995-06-13  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 2  
LENGTH: 5361  
TYPE: DNA  
ORGANISM: P. falciparum  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1)..(5361)  
US-08-973-462-2

## Query Match

Best Local Similarity 45.9%; Score 56; DB 3; Length 5361;  
Pred. No. 0.003; Mismatches 225; Indels 0; Gaps 0;

Matches 191; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

Qy 3869 ACTGTTGCTAAAGTTATGAGATGAAAAATCTCTCTATTAATCAATGATGATCAT 3928  
Db 289 AATGATGATTAAGCTAGTGAAGAGTAAAGAAAAATCTTGACTTATTAAGAGAGGA 348  
Qy 3929 ACTGAATCTGACAGACTCGGAAATATGATGCAATTCAGAGATCGCATTCATCG 3988

Db 349 AATACATTACTGAAGTGTAGATGATATATTAATTAAGAGACCGGAGATATTAAG 408  
Qy 3989 AAAATGAAATATATCACCAGTATTATGAGAAAAATATCTGAATATCAAAAAACAT 4048  
Db 409 GAAATATCTTATTAAGTATATAGAAAGAACCAAAATATATTTGACAAATTTATTA 468  
Qy 4049 GAAATATCTTCTGATTAAGAGTTGATTAATTTGAAACGTTATATTTGATGAAT 4108  
Db 469 AATATATTTGACAAATATCTGAAAAAACAAAGTGTATCAGAAAAATGTCAGT 528  
Qy 4109 GAGAAATGATTAATGACGATGATCTGTAAGCTATCCATACCAAGATGAATATAT 4168  
Db 529 GATGACTTTTATTAATGATTTATTAATGATGATGATGATGAGAGAGTAAAGAAAT 588  
Qy 4169 GATGTTGATGAATGAGAGTTTGTATATATCATGATGACAGACCAAGAAAGAA 4228  
Db 589 ATTTGAGAGAAAGTCAAGTTAATGAGATTTTATATGTTTATGTAAGAAAGTTTCAA 648  
Qy 4229 TACGATTTGAGAAAGATTAAGAAAAATGCTGTCTTAACATTCCTGGA 4284  
Db 649 CAAGAACACACACATGTTGAAAGAAAGTTGAAGAACTGTAGAGAAATGA 704

## RESULT 8

US-08-973-462-1  
Sequence 1, Application US/08973462B

Patent No. 6191270  
GENERAL INFORMATION:  
APPLICANT: DRUILHE, PIERRE  
APPLICANT: DAUBERSIES, PIERRE  
TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
FILE REFERENCE: 0660-0125-0 PCT  
CURRENT APPLICATION NUMBER: US/08/973,462B  
CURRENT FILING DATE: 1998-02-06  
EARLIER APPLICATION NUMBER: PCT/FR96/00894  
EARLIER FILING DATE: 1996-06-12  
EARLIER APPLICATION NUMBER: FR 95/07007  
EARLIER FILING DATE: 1995-06-13  
NUMBER OF SEQ ID NOS: 29  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 6152  
TYPE: DNA  
ORGANISM: P. falciparum  
US-08-973-462-1

Query Match  
Best Local Similarity 45.9%; Score 56; DB 3; Length 6152;  
Pred. No. 0.0031; Mismatches 225; Indels 0; Gaps 0;

Matches 191; Conservative 0; Mismatches 225; Indels 0; Gaps 0;

Qy 3869 ACTGTTGCTAAAGTTATGAGATGAAAAATCTCTCTATTAATCAATGATGATCAT 3928  
Db 533 AATGATGATTAAGCTAGTGAAGAGTAAAGAAAAATCTTGACTTATTAAGAGAGGA 592  
Qy 3929 ACTGAATCTGACAGACTCGGAAATATGATGCAATTCAGAGATCGCATTCG 3988  
Db 593 AATACATTAACTGAAGTGTAGATGATATTAATTAAGAGAGCCGAGATATTAAG 652  
Qy 3989 AAAATGAAATATATCACCAGTATTATGAGAAAAATATCTGAATATCAAAAAACAT 4048  
Db 653 GAAATATCTTATTAAGTATATAGAAAGAACCAAAATATATTTGACAAATTTATTA 712  
Qy 4049 GAAATATCTTCTGATTAAGAGTTGATTAATTTGAAACGTTATATTTGATGAAT 4108  
Db 713 AATATATTTGACAAATATCTGAAAAAACAAAGTGTATCAGAAAAATGTCAGT 772  
Qy 4109 GAGAAATGATTAATGACGATGATCTGTAAGCTATCCATACCAAGATGAATATAT 4168  
Db 773 GATGACTTTTATTAAGTATATAGTATAGTATGATGAGAGAGTAAAGAAAT 832  
Qy 4169 GATGTTGATGAATGAGAGTTTGTATATATCATGATGACAGACCAAGAAAGAA 4228  
Db 833 ATTTGAGAGAAAGTCAAGTTATGACGATATTTTATATGTTTATGTAAGAAAGTTTCAA 892



QY 4229 TACACATTGAGAAAGAGTTTACGAAAAAATGCTGCTACTTAACATTCCTGA 4284  
DB 893 CAAGAACACACACACATGTTGAGAAAAAGTTGAGAAAGTTGAGAAATGA 948

## RESULT 9

US-09-640-173-53/c  
Sequence 53, Application US/09640173  
Patent No. 6613515  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Stolk, John A.  
TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND  
FILE OF INVENTION: METHODS OF USE THEREFOR  
FILE REFERENCE: 210121.484C2  
CURRENT APPLICATION NUMBER: US/09/640.173  
CURRENT FILING DATE: 2000-08-15  
NUMBER OF SEQ ID NOS: 196  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 53  
LENGTH: 396  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(396)  
OTHER INFORMATION: n = A,T,C or G  
US-09-640-173-53

Query Match 0.9%; Score 55.6; DB 4; Length 396;  
Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1130 GAAGAAATGTGATCATCTGATATGCTAGATTGTTGATCCAGCAACAATAT 1189  
DB 304 GAANNATNN 245  
QY 1190 GAACCTATCTTATGTTAGTTTACACACAGAGAAACCAAAAGAAAACGAGAG 1249  
DB 244 AAGGAAAAAANNN 185  
QY 1250 AACTCAGTGAACATCTCAAGAGAAAGAACTGAATCAAGAAAGAAATAGAAAT 1309  
DB 184 AA 125  
QY 1310 CCAAAATCAGTAAAGATTAAGGTAAGAAAAAGAAAAAGAAAAAATTCCTG 1369  
DB 124 AA 65  
QY 1370 TGAATAACA 1378  
DB 64 AAAAAAAAAA 56

## RESULT 10

US-09-713-550-53/c  
Sequence 53, Application US/09713550  
Patent No. 6617109  
GENERAL INFORMATION:  
APPLICANT: Xu, Jiangchun  
APPLICANT: Stolk, John A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
FILE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.484C4  
CURRENT APPLICATION NUMBER: US/09/713.550  
CURRENT FILING DATE: 2000-11-14  
NUMBER OF SEQ ID NOS: 205  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 53  
LENGTH: 396  
TYPE: DNA  
ORGANISM: Homo sapien

FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(396)  
OTHER INFORMATION: n = A,T,C or G  
US-09-713-550-53

Query Match 0.9%; Score 55.6; DB 4; Length 396;  
Best Local Similarity 47.4%; Pred. No. 0.0014;

Matches 118; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 1130 GAAGAAATGTGATCATCTGATATGCTAGATTGTTGATCCAGCAACAATAT 1189  
DB 304 GAANNATNN 245  
QY 1190 GAACCTATCTTATGTTAGTTTACACACAGAGAAACCAAAAGAAAACGAGAG 1249  
DB 244 AAGGAAAAAANNN 185  
QY 1250 AACTCAGTGAACATCTCAAGAGAAAGAACTGAATCAAGAAAGAAATAGAAAT 1309  
DB 184 AA 125  
QY 1310 CCAAAATCAGTAAAGATTAAGGTAAGAAAAAGAAAAAGAAAAAATTCCTG 1369  
DB 124 AA 65  
QY 1370 TGAATAACA 1378  
DB 64 AAAAAAAAAA 56

## RESULT 11

US-09-410-464-9  
Sequence 9, Application US/09410464  
Patent No. 6395892  
GENERAL INFORMATION:  
APPLICANT: Strauss et al.  
TITLE OF INVENTION: Floral homeotic genes for manipulation of flowering in  
FILE REFERENCE: 53375  
CURRENT APPLICATION NUMBER: US/09/410.464  
CURRENT FILING DATE: 1999-10-01  
EARLIER APPLICATION NUMBER: 09/287,700  
EARLIER FILING DATE: 1999-04-06  
EARLIER APPLICATION NUMBER: 60/080,851  
NUMBER OF SEQ ID NOS: 24  
SOFTWARE: Patent Ver. 2.0  
SEQ ID NO 9  
LENGTH: 11485  
TYPE: DNA  
ORGANISM: Populus balsamifera subsp. trichocarpa  
US-09-410-464-9

Query Match 0.9%; Score 55.6; DB 4; Length 11485;  
Best Local Similarity 44.6%; Pred. No. 0.0049;

Matches 348; Conservative 0; Mismatches 424; Indels 8; Gaps 3;

QY 4468 CAACAAATCGACTTATGATCCGGAGTGAATGAATGAACTGAACTAGAGA 4527  
DB 130 CTATAAGAGGCAATCTTGATCCGCTCGGCAACAACCTAATGAATGAATTTCTGC 189  
QY 4528 ATTAGATCAAAAGATGTTTACGAGAAGTTTCAATTCACACGGTGAAGCCATATC 4587  
DB 190 TTGCAATAGACAGTACTTGGAATCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCT 249  
QY 4588 TATGGGTTGGTACATCTGAGAAATGATTTCTCTCAAGGTTGTTGGAAATCAG 4647  
DB 250 TTGCAATGGGTTGATTAAGATCAAGCTAATCT--GATGGTCTATTAAGCAATACA 306  
QY 4648 TTGCTGTCATGAGCAAGCAAGAAATGGAATTAAGCCCTTTAGGTGAG 4707  
DB 307 AGCTAGCGCTTGGCAAAAGATCTCTCAACATTATGATGAGTATGAGAAACATT 366

QY 4708 TTCACTGTTATAGATCTTGTGACATTAATGACAAATAATAGTGTGATTAAG 4767  
DB 367 TGGCCCGTTGCAAAAATGACTACTATCTTACTTATTTGTGAGCTTGCATTTGCA 426  
QY 4768 AATGACATTCACATTTAGACGTGAGTGGGTAATCTAATGCTTATTAATCTC 4827  
DB 427 GTGCAATTTCTCAGCTGTATGTAATAATGCTTGTGAATGAGATCTTCAAGAGA 486  
QY 4828 AATTCATTTATGTCCTTCTCTCTAATGATACCTTTGAAGAAAACATTTGGT 4887  
DB 487 AGTTATGTCGACCTCCCTCTGTAATTCATATGACTC--TGGATATGTTTGAACCTT 544  
QY 4888 ATTGAACGTTCTGATGAGGTTTAAACAGTCGGGTTTGAATGGTATCACAATCA 4947  
DB 545 AAGAAAGCATTAATTAATGCTCTCAACAGACACCCCGCTGTGTTGAGAAATTC 604  
QY 4948 AAGAGATTTGGAACATTTGCTTTACTCAAGTTTACACATGATGTTTATTTCACT 5007  
DB 605 TATGTGATCTCGCTCTCTGCAATTTGTTCTAGACATGATTCGCTCTTTT--AT 661  
QY 5008 TGAATATGAAGAGGATCAGTAATATATTAGTTTATGTTGATGATCTTATGCT 5067  
DB 662 TAAATGACATGATGACAGTGTGATCATTCTCTTATATGTTGATTAACATGATTTAT 721  
QY 5068 TGAAGTTCACAAAAGTTATGATAATTTTGTGATCAATGAGATCATTGTAAGT 5127  
DB 722 TGGTATGACATGATGATGATTTCACTTGAAGACAAAGTTGGCTAGACATTTGAAT 781  
QY 5128 TAAAGTGTGTTGTAATTAATCAATTTCTGATTAATTTGTAATTTGTAATTTG 5187  
DB 782 GAAAGATTTGGTTATCTTCAATATTTCTGAGGTAATGAGTACATCTCACCCTAGAG 841  
QY 5188 TTATATTTTATCTCAAGAAAATTTCTCAAGAAATTAATTAAGATTTCAACTAGTA 5247  
DB 842 TTACTCTCTCTCTGAGTGAATAATGTTGAGATATTCTGAGACAGCTAGACTTACTGA 901

## RESULT 12

US-09-254-776B-50

Sequence 50, Application US/09254776B

Patent No. 6559359

GENERAL INFORMATION:

APPLICANT: late, Howard

TITLE OF INVENTION: PLANT RETROVIRAL POLYNUCLEOTIDES AND METHODS FOR USE THEREOF

FILE REFERENCE: 27013/33479A

CURRENT APPLICATION NUMBER: US/09/254, 776B

CURRENT FILING DATE: 1999-03-09

NUMBER OF SEQ ID NOS: 86

SOFTWARE: PatentIn version 3.0

SEQ ID NO 50

LENGTH: 2826

TYPE: DNA

ORGANISM: Glycine max

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: SIRE-1 genomic clone

US-09-254-776B-50

Query Match

Best Local Similarity 43.7%; Pred. No. 0.0032;

Matches 343; Conservative 0; Mismatches 436; Indels 6; Gaps 2;

QY 4484 AATGATCCGGGATGATTAAGTCATGATGTAAGTGAAGAAATTTAGTCAAAAGAT 4543  
DB 1324 ACTGATGATTTCTGATCAATGCTATGCAAGAAATTTGAGCAATTCAAAGAAATGA 1383  
QY 4544 GTTTAAGAGAGATTTCAATTTCCACCGGTGAAGCTATATCTATGAGTGTGATCAT 4603  
DB 1384 GTTTGGAGCTAGTTCTCTAGCCCGAGGAACTATATGATTTGACACCACTGATCTTC 1443  
QY 4604 ACTGAGAAATTTGATTTCTCTCAAGTGTGTTGGAATACCGTGTGTTGTCATGAGGC 4663

DB 1444 AAGAAACAAACAAAT--GAAAGAGTGTATATACCAAGAAACAAGCCAGACTTGTGCT 1500  
QY 4664 AACAGACAAAAGAAAATTTGATTTATGACCTTTTATGTTAGTTACCTGTTATATGAT 4723  
DB 1501 CAAGGCTACACTGATTTGAGAGGTGATGACTTTGATGAACCTTTGCCCCGTGGTCAAA 1560  
QY 4724 CTGTGACTATTAATTTATGACAAATATATGTTGTAATTTAGGAATGACAAATTCACAT 4783  
DB 1561 CTGAGTCCATCAACACTGTTACTGTTGTGCTTGATCTCTCAATTTCAAGCTGTACAG 1620  
QY 4784 TTAGACGTGAGTGGGATCTAATGATGCTTATTAATCTATCAATTCAAATTTATGTC 4843  
DB 1621 ATGATGTTGAAGACGATTTCTGATGATGATGATGATGATGATGATGATGATGATG 1680  
QY 4844 TTTCTCTTAATATGATGATCTTTGAAGAAAACATTTGTTATTTGAACGTTCTGTC 4903  
DB 1681 CCAAGGATTTGATGATCTCAACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1740  
QY 4904 TATGGTTTAAACAGTGGGTTTGAATGATGATCACTATCAAAAGATTTGGAAGAC 4963  
DB 1741 TATGATTTGAAGCAAGCTTCAAGAGCTTGTGATGAAGGCTACAGAGTCTTACTAG 1800  
QY 4964 ATGCTTTACTCAAGTTTATACATGATGCTTTATTTCACTGATTAATGAAGAGGA 5023  
DB 1801 CAAGGCTATGAGAAAGGGGGGATGACAGACCTTTTGTG--TTAAACAGATGCTGA 1857  
QY 5024 TCAGTAATATTTATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 5083  
DB 1858 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1917  
QY 5084 GTTATGATTAATTTTGTGATCAATTTGATGATGATGATGATGATGATGATGATGATGATG 5143  
DB 1918 ATGCTTCACTTTTGTGATCACTCACTCACTCACTCACTCACTCACTCACTCACTCACT 1977  
QY 5144 ATATCAATTTATCTGATTTGATGATTTGATGATTTGATGATTTGATGATTTGATGAT 5203  
DB 1978 CTGATTTATTTTGTGGAATCAAGTGAAGCAATGGAATGATGATGATGATGATGATGATGAT 2037  
QY 5204 GAAATATTTCTCAAGAAATTTACTTAAGGATTTCAACTAGATGATGATGATGATGATGATG 5263  
DB 2038 AGCAATGATGCAAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2097  
QY 5264 ATACC 5268  
DB 2098 ACACC 2102

## RESULT 13

US-09-835-811-1

Sequence 1, Application US/09835811

Patent No. 6482936

GENERAL INFORMATION:

APPLICANT: HU, Song et al

TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,

NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND

FILE REFERENCE: CL0012228

CURRENT APPLICATION NUMBER: US/09/835, 811

CURRENT FILING DATE: 2001-04-17

NUMBER OF SEQ ID NOS: 5

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1696

LENGTH: 1696

TYPE: DNA

ORGANISM: Human

US-09-835-811-1

Query Match

Best Local Similarity 57.6%; Pred. No. 0.0029;

Matches 99; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 1207 TAGTTTACCAAGCAAGAGAAAACAAAAGAAAACCAAGGAGAACTCACTGGAACATC 1266

Db 1506 TAGATTAAAAA  
Qy 1267 TCAGAGAGAACTGAAATCAGAGAGAAATPAGAAACATCCAAATCAGATTACGA 1326  
Db 1568 AA 1627  
Qy 1367 TAAAGTGAAAAAGAAAAAGAAAAAACTTCACTGAGATGAAAAACA 1378  
Db 1628 AA 1679

RESULT 14  
US-08-771-602D-1  
Sequence 1, Application US/08771602D  
Patent No. 5976795

GENERAL INFORMATION:  
APPLICANT: Voytas, Daniel F.  
APPLICANT: Zou, Sigie  
TITLE OF INVENTION: Retrotransposon and Methods  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle, Suite 201  
CITY: Boulder  
STATE: Colorado  
COUNTRY: USA  
ZIP: 80303

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/771.602D  
FILING DATE: 20-DEC-1996  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/010,869  
FILING DATE: 31-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Feiber, Donna M.  
REGISTRATION NUMBER: 33,878  
REFERENCE/DOCKET NUMBER: 8-96  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Saccharomyces paradoxus  
STRAIN: NREL Y-17217

FEATURE:  
NAME/KEY: CDS  
LOCATION: 1441..6321  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1228..6602  
OTHER INFORMATION: /function= "retrotransposon"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1228..1478  
OTHER INFORMATION: /product= "Ty5-6p"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 6352..6602

OTHER INFORMATION: /function= "3' LTR of Ty5-6p"  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 2852..4827  
OTHER INFORMATION: /function= "integrase region coding  
OTHER INFORMATION: region of Ty5-6p"  
US-08-771-602D-1

Query Match 0.8%; Score 54.6; DB 2; Length 6660;  
Best Local Similarity 45.2%; Pred. No. 0.0067;  
Matches 241; Conservative 0; Mismatches 289; Indels 3; Gaps 1;

Qy 4661 GGCAACAGCAAAAGAAAAATGATATGACCCCTTTAGTGTGATGACCTGTTA 4720  
Db 4981 GCCCAAGACATACCTAAAGAGCTGTATGACTATCAAGAACTTTGACCACTGAT 5040  
Qy 4721 GATCTTGACTATTAAGATTATGACAAATATAGGTTGAAATGAAATGCAATTC 4780  
Db 5041 CGATATGACTGTTAGATTATTTGCGCCCTTGAGCTGCTCAACTATATGATAT 5100  
Qy 4781 CATTAGAGCTGAGTGGCGGCTATCTAATGCTCTATACCTCAATCCAAATTTAT 4840  
Db 5101 CAGATGACGTTGACACCGCGTTCTAACTCAAAATGATAGCCGCTATACGTA 5160  
Qy 4841 GTCTTCTCTCTAAATCAGTACCTTTGAGAAAAACATTTGTTGTTATTTGAAGCT 4900  
Db 5161 CAACACCGGATTTATTAATGAAGTATCCCGACTATGATGGAATATACGCGGT 5220  
Qy 4901 GTCTATGGTTAAACAGCTGGGTTTGAATGTTATCAGATCAATCAAGATTTGCA 4960  
Db 5221 ATGTATGACTAGCAAGCAAGCCCATTAATGAGCAAGCAATATCAATCTTCA 5280  
Qy 4961 GACATGATTTTACCAATTTTACCAATGATGTTTATTCACATGATGATGAGAG 5020  
Db 5281 AGATTGTTTTCGTGATGATGAGGCAAGCTGCTTATTTGCTTCCATCT 5337  
Qy 5021 GGATCAGTAATATTAGGTTTATGATGATGATGATGATGATGATGATGATGAT 5080  
Db 5338 GATGTCCTCATCTACCTGCTTATGATGATGATGATGATGATGATGATGATG 5397  
Qy 5081 AAGTATGATATTTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 5140  
Db 5398 AATATATGACAGGTTTATGACAAATCAAGCAATGATGATGATGATGATGAT 5457  
Qy 5141 GAATATCAATTTATCTGTTATGATGATGATGATGATGATGATGATGATGAT 5193  
Db 5458 AAGTATCAATTTCTGCTTATGATGATGATGATGATGATGATGATGATGAT 5510

RESULT 15  
US-09-232-446B-1  
Sequence 1, Application US/09232446B  
Patent No. 6228647  
GENERAL INFORMATION:  
APPLICANT: Voytas, Daniel F.  
APPLICANT: Gal, Xisewu  
TITLE OF INVENTION: Transposable Element Protein that Directs DNA  
FILE REFERENCE: 2-98  
CURRENT APPLICATION NUMBER: US/09/232,446B  
CURRENT FILING DATE: 1999-01-15  
PRIOR APPLICATION NUMBER: US 60/071,383  
PRIOR FILING DATE: 1998-01-15  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
TYPE: DNA  
LENGTH: 6660  
ORGANISM: Saccharomyces paradoxus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1441)..(6318)  
US-09-232-446B-1



Result No.	Score	Query Match	Length	DB	ID	Description
1	92	1.4	1289	12	US-10-424-559-11662	Sequence 71662, A
2	89.6	1.4	5447	12	US-10-424-559-103057	Sequence 103057, A
3	84.2	1.3	4269	12	US-10-424-559-12495	Sequence 42495, A
4	84.2	1.3	4133	12	US-10-424-559-12491	Sequence 42491, A
5	83.8	1.3	13341	8	US-08-910-386A1	Sequence 1, Appl
6	79.6	1.2	78333	15	US-10-298-112-3	Sequence 3, Appl
7	70.6	1.1	560	12	US-10-424-559-116696	Sequence 11696, A
8	70.6	1.1	2760	12	US-10-424-559-108338	Sequence 10838, A
9	68.4	1.1	3673778	14	US-10-311-455-62411	Sequence 1, Appl
10	68.2	1.1	7306	14	US-10-311-455-610	Sequence 1610, Appl
11	65.8	1.0	3673778	14	US-10-412-441-2	Sequence 2, Appl
12	65.6	1.0	7668	12	US-10-442-559-108640	Sequence 108640, A
13	65.4	1.0	11729	14	US-10-311-455-868	Sequence 868, Appl
14	65.2	1.0	661	9	US-09-822-830A-66	Sequence 26, Appl
15	65.2	1.0	12177	14	US-10-311-455-624	Sequence 624, Appl

16	65	1.0	628	12	US-10-424-599-62662	Sequence 62622, A
17	65	1.0	2685	12	US-10-424-599-108648	Sequence 108648, A
18	64.8	1.0	14798	14	US-10-311-455-1005	Sequence 1005, Ap
19	63.4	1.0		14	US-10-032-585-2429	Sequence 2429, Ap
20	63.4	1.0	9358	15	US-10-396-122-90	Sequence 90, Appl
21	63.4	1.0	24259	12	US-10-221-714-A-415	Sequence 415, Appl
22	63.2	1.0	2190	12	US-10-424-599-9933	Sequence 9933, Ap
23	63.2	1.0	3861	12	US-10-424-599-8892	Sequence 8892, Ap
24	63.2	1.0	3879	12	US-10-424-599-8933	Sequence 8903, Ap
25	63.2	1.0	3912	12	US-10-424-599-8914	Sequence 8914, Ap
26	63.2	1.0	335913	10	US-09-754-853-A-2	Sequence 2, Appl1
27	63.2	1.0	335913	10	US-09-754-853-A-3	Sequence 3, Appl1
28	62.8	1.0	6192	12	US-10-321-714-A-408	Sequence 408, App
29	62.4	1.0	8946	14	US-10-311-455-884	Sequence 884, App
30	62	1.0	767	12	US-10-424-599-45104	Sequence 45104, A
31	62	1.0	9515	14	US-10-239-676-160	Sequence 160, App
32	62	1.0	9515	14	US-10-240-453-182	Sequence 182, App
33	62	1.0	10369	12	US-10-321-714-A-25	Sequence 25, Appl
34	62	1.0	10369	12	US-10-311-455-35	Sequence 35, App
35	61.8	1.0	1460	12	US-10-424-599-14665	Sequence 14625, A
36	61.4	1.0	327	9	US-09-960-352-4630	Sequence 4630, Ap
37	61.4	1.0	799	12	US-10-424-599-88818	Sequence 88818, A
38	61.4	1.0	4985	14	US-10-094-240-10	Sequence 10, Appl
39	61.4	1.0	4985	14	US-10-056-405-10	Sequence 10, Appl
40	61.2	1.0	1198	9	US-09-789-561-55	Sequence 55, Appl
41	61	0.9	1113	14	US-10-083-357-599	Sequence 599, App
42	61	0.9	2973	12	US-10-425-114-13139	Sequence 13139, A
43	61	0.9	2981	12	US-10-424-599-58746	Sequence 58746, A
44	61	0.9	9399	15	US-10-396-122-93	Sequence 93, App
45	60.8	0.9	973	12	US-10-424-599-58679	Sequence 58679, A

## ALIGNMENTS

RESULT 1  
 US-10-424-599-71662  
 Sequence 71662, Application US/10424599  
 Publication No. US20040031072A1  
 GENERAL INFORMATION:  
 APPLICANT: La Rosa Thomas J  
 APPLICANT: Kovalic David K  
 APPLICANT: Zhou Yihua  
 APPLICANT: Cao Yongwei  
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53223)B  
 CURRENT APPLICATION NUMBER: US/10/424,599  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 285684  
 SEQ ID NO 71662  
 LENGTH: 1289  
 TYPE: DNA  
 ORGANISM: Glycine max  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: (1)..(1289)  
 OTHER INFORMATION: unsure at all n locations  
 FEATURE:  
 OTHER INFORMATION: Clone ID: PAT\_MRT3847\_35724C.1  
 US-10-424-599-71662

Query Match	Similarity	1.4%	Score 92	DB 12	Length 1289
Best Local	Similarity	46.3%	Pred. No. 5.5e-08		
Matches	376	Conservative	0	Mismatches 430	Indels 6
				Gaps	2
QY	4486	TATATCCGGATGATGATTAAGTATGCTGCAATAGAGAAATTTAGATCAAAAGTCT	4545		
DB	297	TGATCAAAATGGAATATGAAATGAAAGAGAGCTTTATATGATGAAAAATTTAAAC	356		
QY	4546	TACGAGAGAGTTCCATTTCCACCGGTGGAAGCTTATCTATGAGTTGGGTACATTC	4603		
DB	357	ATGGGAGCTGTGTGACACAGCTTAACCAAGAAAGCATTTGGTGTGAAGTGGTTATAG	416		

```

QY 4606 TGAGAAATGATCTCTCAAAAGGTGTGTCGAAATACGCTTGTGTCCATGCGCA 4665
DB 417 AACCAAGCTCAATTC---AGATGGTCTATTAACAATAACAGCAAGGCTTGTGTAA 473
QY 4666 CAGCAAAAAGAAAAATTTGATTTTGAACCTTTTACGTAGTTCACCGTCTATATCT 4725
DB 474 AGGATATGCTCAAGTGTGTGTGTATTTTCAGAGACTTTTGCCTTACCTCCAGACT 533
QY 4726 TGTACTATATAGATTTATTTGACAAATATAGTGTGATTTAGATGACATTCACATTT 4785
DB 534 AGAAACCTTAACACAGGGGCGGTCTGTGTGACCAAAAAGGTGATTTATCATCAAT 593
QY 4786 AGAGCTGAGTCGGCGATCTAAATGCTCTATTTACTTCAATTCATTTATGTCTT 4845
DB 594 GGATGTGAATACACCTCTTAAATGATCTTGAGAGAAATTTTGTGACACACC 653
QY 4846 TCCCTTAATCACTACCTTTGAGAGAAAAACATTTGTGTTATTTGAAACGTTCTGCTA 4905
DB 654 TGAAGGTTGCTGTGTTCAGAGACAGAGAAAGTATTCGATTAAGAAAGCTTTGTA 713
QY 4906 TGGGTTAAAACAGTCGGGTTTGAATGTGATCACACTATCAAAAAGATTTGAAACAT 4965
DB 714 TGGTTAAAGCAGCGCACCGCGTCTATGATGAGAGATTTGATGACACTTGTGAACTT 773
QY 4966 TGGTTTACTCAAGTTTATACCAATGATGGTTTATTTTCACATTGAATGAGAGGATC 5025
DB 774 AGGCTTCAAAAAGTTTAACTGATCTACTTTATATGTAAGAACTAATGTGAAT 833
QY 5026 AGTAAATATTTAGGTTTAAATGTTGATGATATTTCTTATGTTGAAGTTTACAAAAAGT 5085
DB 834 GCTTGTAA---GTATCTCATATGTTGATGATCTACTGTTTCGGGAAGCAGCGGGAGCA 890
QY 5086 TATGATATTTTGTGATCAATTGAGAGATCTTTGAGATTGAAGTTTGTGTAAT 5145
DB 891 TATGACAAAGTTCAGAGAAAGAAAGAGATGCTTTGAAATGACACCTTGAAGAT 950
QY 5146 ATCAATATCTTGTGATTTGATTTTGTAAAACCGAATCTGTTATATTTATCTCAGA 5205
DB 951 GACATCTTCTTCGTGATGAGGTGATCAAAAACAAAGAAATATTTCTTGGCAGGA 1010
QY 5206 AAAATTTCTCAAGAAATTTACTTAAAGATTTCAAACTGATGACATATGGAAGAAAT 5265
DB 1011 AAAATATGCAATGAAATTTCTCAAGAGTTTCACATGGAATGTGCAAAACCACTGCAC 1070
QY 5266 ACCCGATTCGATGACAAATATGAAAAGG 5297
DB 1071 TCCATGATCAAAAAGAGAGTTTACAAAG 1102

```

```

RESULT 2
US-10-424-599-103057
; Sequence 103057, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 103057
; LENGTH: 5417
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_64078C.1
US-10-424-599-103057

```

```

Query Match 1.4%; Score 89.8; DB 12; Length 5417;
Best Local Similarity 46.8%; Pred. No. 3e-07;
Matches 283; Conservative 0; Mismatches 322; Indels 0; Gaps 0;
QY 4664 AACAGACAAAAGAAAAATTTGATTTATGACCCCTTTTATGTTTACTTACCTGTTATAGAT 4723
DB 3760 AAAGGTTATAGCAACATATGAGTGTATATGATGAGTGTGACCGGTTGCCCGC 3819
QY 4724 CTTGATCTATTAAGTTTATTCACATTAATAGTGTGATTTAGAAATGACATTCACAT 4783
DB 3820 ATGAAACCATTCGCTCTTATTTCTTGGACCTCAAAAGAGTGAATTTTTCAG 3879
QY 4784 TTAGAGTCAGTCGGCGTATCTAAATGCTCTATTTACTATTCATTCATTCATTTATGTC 4843
DB 3880 CTGATGTAATATGCGCATTTCTAAATGCGTATCTTGAAGAAATGCTATGTTGAACAA 3939
QY 4844 TTTCCTCTTAATCAGTACCTTTGAGAGAAAAACATTTGTGTTATTTGAAACGTTCTGTC 4903
DB 3940 CCAATGGGTTTGTATGAGAGGTCAGAAAGAAAGTCTGAAATTTGAACAGCGCTTG 3999
QY 4904 TATGGTTTAAAACAGTCGGGTTTGGAAATGATCACTATCAAAAAGTATTTGAAAGAC 4963
DB 4000 TATGTTTAAAGCAGACCGAGGCAATGCAATCTGCAATGACAGTATCTTCCAGAC 4059
QY 4964 ATGTTTACTCAAGTTTACACATGATGTTTATTTTCACTGATTAATGAAAGAGGA 5023
DB 4060 AATGGGTTTGTGCTGTCAAAATGAGTATCTCTTATGTTAAACCTTTTAAATATGCT 4119
QY 5024 TCAGTAATATTTAGGTTTATTTATGTTGATGATATTTCTTATGTTGAGTTCACAAAA 5083
DB 4120 GATGCTTATTTATTTGTTTATTTATGATGATACCTTATCTTACCAGCAATACCCAAAT 4179
QY 5084 GTTATGATTAATTTTGTGATCAATTTGAGATCAATTTTGAAGTTAAAGTTTGTGTA 5143
DB 4180 TTGTTTAAAGCTTCAAGAGTCCATGCTGTAATTTGATGACAGATATGAGGACTC 4239
QY 5144 ATATCAATATCTTGTGATTTGATTTTGTAAAACCGAATCTGTTATTTATTTATCTCA 5203
DB 4240 ATGCTATTTACTTGGGATGGAAGTGAAGCAACCGGAATGGAATCTTGTCTCACA 4259
QY 5204 GAAATTTCTCAAGAAATTTACTTAAAGATTTCAAACTGATGACATATGGAAGAAAT 5263
DB 4300 GAAAGTACAAAGAAAGTGTGAAGAAATTTATATGTTGATGCAATCCCGTGAAC 4359
QY 5264 ATACC 5268
DB 4360 ACACC 4364

```

```

RESULT 3
US-10-424-599-42489
; Sequence 42489, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 42489
; LENGTH: 4029
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_138369C.1
US-10-424-599-42489
Query Match 1.3%; Score 84.2; DB 12; Length 4029;

```



Query Match	1.3%	Score 83.8	DB 8	Length 13341
Best Local Similarity	46.4%	Pred. No. 7.1e-06		
Matches 345	Conservative	0	Mismatches 392	Indels 6
				Gaps 2
Qy	4496	TCGATTAAGATCAATGATAGCTGAACCTAGAGAAATTTAGATCAAAAAGATGTTTACGAGAA	4555	
Db	7328	TGGAGAGATGCAATGGAAGAACTGAGTATTAAGCTTTGATTAAGAAAAGACACATGGCACCCTA	7387	
Qy	4556	GTTTCAATTCGCCACCGGGTGGGAAGCCATATCTATGAGTGGGTGATCATCTAGAAAATT	4615	
Db	7388	GTTTCAATTTGGAAGAAAGCAAAATATCATTTGGGTGTAATGGGTATATTAAGTTAAAG	7447	
Qy	4616	GATTCTCTCAAAAGGTGTTGTCGAAATCACGTTGTGTTGCCATGCGAACAGCAAAAG	4675	
Db	7448	AAGCAGATGGGACACTTGATAGATCAAAAGCTAGACTGT--AGCAAGGGCTTTTAA	7504	
Qy	4676	GAAAAATTGATTTAGAACCCCTTTAGGTAGTTCACCTGTATAGATCTTGTGACTATA	4735	
Db	7505	CAAAATATGATATGATATGATTAAGAAATCTTTAGTCCTGTTGTTAAGCTGCTACTATT	7564	
Qy	4736	AGATTAATTGACAAATATAGGTGTGGAATTGGAATGACAAATTCACATTAGACGTGAG	4795	
Db	7565	AGAAATTTATGTCATATGCTGATGCTGCTCTAGAGGTGAGGCTTAGACAGTATATGTTGAG	7624	
Qy	4796	TGGGGTATCTAAATTAAGCCTCATATCTACATCAATCAATTCATTAATGCTCTTCTCTTAA	4855	
Db	7625	AATGCTTTTCTTCATGATTTTGAAGAAAGTCTACATGCAACACCTCTCTGCTTT	7684	
Qy	4856	TCAGTACCTTTGAGAAAAAACCAATTGTGGTATGTAAACGTTCTGTCTATGGTTAAA	4915	
Db	7685	GAGTCATCTCTTAAACCTGATTAATGATGTAAATGTGATAGGCAATATATGCGCTGAAA	7744	
Qy	4916	CAGTCGGGTTTGAATGCTATACACATCAAAAGGTATGGAAGACATGCTGTTTACT	4975	
Db	7745	CAGACACCAAGGCGGTGATTCAGGCTAGTAAAGAACTGTTGAACCTGTTTT--T	7801	
Qy	4976	CAGTTTATACAAATGATGTTTATTTTACATTGAATTAAGAGGATCACTAATATAT	5035	
Db	7802	GAACTTCAAAAGCGCTGATACCTCATATTTCTTTCTTAAACAAAGAGGGAATCTTATGTTT	7861	
Qy	5036	TTAGTTTATATGTTGATGATATCTATGTTGGAAGTTCACAAAAGTATTTATCATAT	5095	
Db	7862	GTTTGTGATATGTTGATGATATATGTTAGCTCTACAGAGAGGCACTACAGCA	7921	
Qy	5096	TTTGTGATTCATTCAGATCATTTTGAAGTTAAAGTGTGGTGAATATCAATATAT	5155	
Db	7922	CTTCTGAAGGATCTTAAACAGAGATTCGCACTTAAGATTTGGGAGACCTGCACATCTTC	7981	
Qy	5156	CTTGTATTTGATTTTCGTAACCCGAATCTGTATATTTTATCTCAAAAATTTCTC	5215	
Db	7982	CTTGAATTTGAGTACTTAAAGTTTCCAAATGGCGTTATCTTGACTCAGAGATATGCA	8041	
Qy	5216	AAGAAATTACTTAAGGATTTCAA	5238	
Db	8042	AATGATCTGCTTAAAGAGATTAA	8064	
RESULT 6				
US-10-298-122-3/c				
Sequence 3, Application US/10298122				
Publication No. US20030221214A1				
GENERAL INFORMATION:				
APPLICANT: Giltner, Frederick G				
APPLICANT: Deng, Zhao				
APPLICANT: Zhang, Hongbin				
TITLE OF INVENTION: CITRUS TRISTEZA VIRUS AND METHODS OF USE				
FILE REFERENCE: 5853-220				
CURRENT APPLICATION NUMBER: US/10/298.122				
NUMBER OF SEQ ID NOS: 19				
SOFTWARE: Patent version 3.1				
SEQ ID NO 3				



LENGTH: 78333  
 TYPE: DNA  
 ORGANISM: artificial  
 FEATURE:  
 OTHER INFORMATION: Poncirus trifoliata and citrus grandis hybrid  
 US-10-298-122-3

Query Match 1.2%; Score 79.6; DB 15; Length 78333;  
 Best Local Similarity 46.1%; Pred. No. 0.00011;  
 Matches 385; Conservative 0; Mismatches 439; Indels 12; Gaps 3;

QY 4462 ATATGATCCGGGATGCTAAAGTCATGAACTGTAAGTGAATAATTGATCAAAAG 4541  
 DB 44529 AAAATGATGAATCTTGATTTGCAAGTGAAGAGATGATCAATTTGAGAAATA 44470  
 QY 44542 ATGTTACGAAGAAGTCCCAATCCCGGTGGAAGCCATATCTATGGGTGGTAC 4601  
 DB 44469 ATGCTGGAGATGATCTTATCCGGAATATTAATCATTAAGTAAATGGGTAT 44410  
 QY 4602 ATACTGAGAAATGATCTCTCAAGGTGTTGCG--AAATCAGTTGTTGTC 4658  
 DB 44409 TTAGAAACAAATGATGAATCCGGTGTGTTGAATAAAGCTAGATTAGTGCCTC 44350  
 QY 4659 ATGCAACAGCAAAAGGAAATTTGATGACCTTTAGTGTAGTTCACTGTGA 4718  
 DB 44349 AAGGTACACCAAGAGAGAAATGATTTGATGAACCTTTG-----CACCTGAG 44296  
 QY 4719 TAGATCTGTGACTAATGATTAATGACATATAGTTGATGATGAAATGCAATTC 4778  
 DB 44235 CAAGTGTGAATCCATGATGATGATGATGATGATGATGATGATGATGATGAT 44236  
 QY 4779 AACATTGAGGTGAGTGGCGGTATCTAATGCTTATTAATCAATCAATTT 4838  
 DB 44235 ATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 44176  
 QY 4839 ATGCTTTCCTCCTAATCAGTACCTTTGAAGAAACCATGTTGTTATGAAAGCT 4898  
 DB 44175 AACAACTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 44116  
 QY 4899 CTGTCTATGGTTAAACAGTGGTGTGATGATGATGATGATGATGATGATGATGAT 4958  
 DB 44115 CTGTCTATGGTTAAACAGTGGTGTGATGATGATGATGATGATGATGATGATGAT 44056  
 QY 4959 AAGATATGTTTACTCAAGTTTACAGTATGATGATGATGATGATGATGATGATGAT 5018  
 DB 44055 TGGATATGATTTTTC--AAATGGAAAGCGGACACACTCTTTTGTGATGATGAT 43939  
 QY 5019 AGGATCAGTATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 5078  
 DB 43998 ATCAAGACATCTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 43939  
 QY 5079 AAAAGTATGATATTTTGTGATCAATGAGATCAATTTGAACTTAACTGTTG 5138  
 DB 43938 AGTTGTTGTAAGATTTTTCATCATGATGATGATGATGATGATGATGATGATGAT 43879  
 QY 5139 GTGAATATCAAAATATCTGTATGATGATGATGATGATGATGATGATGATGATGAT 5198  
 DB 43878 GAGATGATGATTTCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 43819  
 QY 5199 CTCAGAAATTTTCTCAAGAAATTAATTAAGATTTCAAACTGATGATCATATGAGGA 5258  
 DB 43818 ACCAAGCCAGTACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43759  
 QY 5259 AAAACATCCCGGATCCGATGATGATGATGATGATGATGATGATGATGATGATGAT 5314  
 DB 43758 AAGGATCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43703

RESULT 7  
 US-10-424-599-118696  
 ; Sequence 118696, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:

APPLICANT: La Rosa Thomas J  
 APPLICANT: Kovalic David K  
 APPLICANT: Zhou Yihua  
 APPLICANT: Cao Yongwei  
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53223)B  
 CURRENT APPLICATION NUMBER: US/10/424,599  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 285684  
 SEQ ID NO 118696  
 LENGTH: 560  
 TYPE: DNA  
 ORGANISM: Glycine max  
 FEATURE:  
 OTHER INFORMATION: Clone ID: PAT\_MRT3847\_78191C.1  
 US-10-424-599-118696

Query Match 1.1%; Score 70.8; DB 12; Length 560;  
 Best Local Similarity 51.7%; Pred. No. 0.00057;  
 Matches 187; Conservative 0; Mismatches 172; Indels 3; Gaps 1;

QY 4868 AAGAAACCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 4927  
 DB 45 AAGCCATCATGCTCTTAATCACTGAAAAAGTTTATATGATGATGATGATGATGATGAT 104  
 QY 4928 GAATGTTACACTATCAAAAGAGTATGAAAGCATTTGTTTACTCAAGTTTACAC 4987  
 DB 105 GCTTGTATGAACCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 161  
 QY 4988 AATGATGTTTATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5047  
 DB 162 GTTATCTACCTTTTATTTTAAAGAAATGATGATGATGATGATGATGATGATGAT 221  
 QY 5048 GTTATGATATTTCTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5107  
 DB 222 GTTATGATATTTATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 281  
 QY 5108 TTGAGATCATTTTGAAGTTAAAGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 5167  
 DB 282 ATGCAAAATGAATTTGAAATGCAATGATGATGATGATGATGATGATGATGATGATGAT 341  
 QY 5168 TTTCGTTAAACCAATCTGTTATATTTATCTCAAGAAATTTCTCAAGAAATTTACTT 5227  
 DB 342 ATCAAGCAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 401  
 QY 5228 AA 5229  
 DB 402 CA 403

RESULT 8  
 US-10-424-599-108338/C  
 ; Sequence 108338, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:  
 APPLICANT: La Rosa Thomas J  
 APPLICANT: Kovalic David K  
 APPLICANT: Zhou Yihua  
 APPLICANT: Cao Yongwei  
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
 FILE REFERENCE: 38-21(53223)B  
 CURRENT APPLICATION NUMBER: US/10/424,599  
 CURRENT FILING DATE: 2003-04-28  
 NUMBER OF SEQ ID NOS: 285684  
 SEQ ID NO 108338  
 LENGTH: 2760  
 TYPE: DNA  
 ORGANISM: Glycine max  
 FEATURE:  
 OTHER INFORMATION: Clone ID: PAT\_MRT3847\_68844C.1  
 US-10-424-599-108338

Query Match 1.1%; Score 70.6; DB 12; Length 2760;  
 Best Local Similarity 45.3%; Pred. No. 0.0013;  
 Matches 367; Conservative 0; Mismatches 434; Indels 6; Gaps 3;

```

Qy 4496 TGGATTAAGTCATGTAATGCTGTAAGTGAAGAAATTTAGATCAAAAGATGTTTACGAAGAA 4555
Db 1503 TGGATCAATTCGCAATGCAAGAAAGAACTAAACCAATTTGAAAGAAACAAATGATGAAATTA 1444
Qy 4556 GTTCCATTTCCCAACCGGTGAGACCTATATCTATGSGTGGGTGACATCTGAGAAATTT 4615
Db 1443 GTAGAAAAAAGCTGAAAAATTTCTCTGTTATGGAACAAAAGGGGTTTTTAAATAATTA 1384
Qy 4616 GATTCCTGCAAGAGTGTGTTGCGAAATTCAGTTGTGTGTCATGCGACACAGAAAAAG 4675
Db 1383 GATGAACATGTAATTAATTTAGAAATTAAGCCAGTTAGT---AGAAAAAGGTATTAAT 1327
Qy 4676 GAAAAATTTGATTAAGACCTTTTATGTTAGTTCACCTGTTATAGATCTTTGACTATA 4735
Db 1326 CAAGAGAGGAGATAGACTATGAAAGAAACATATGCTGTTGCAAGATTGAGAGCATTT 1267
Qy 4736 AGATTATGACATAATAGGTGTAATGTAATGGAATGACATTCACATTTAGACGTGAG 4795
Db 1266 AGAATGCTTTGGCATATGATCCATATGAACTTTAACTTTATCAAAATGATGTTAG 1207
Qy 4796 TGGCGATCTTAAATGCTCTATTACTCATTCATTCAAATCCAAATTAATGCTTTCTCTTAA 4855
Db 1206 AGTGCTTTTAAATGCTTAAATCAAGAGAGTATGTTGAAACAAACCCCTGTTT 1147
Qy 4856 TCAGTACCTTTGAAGAAAAACCATTTGTTGTTATGAAAGTTCTGCTATGSGTTAAA 4915
Db 1146 GAAATTTCTGATTAACCAACCAATGTTATTAATTAACAAAGGCTCTTTATGTTGAAA 1087
Qy 4916 CAGTCGGATTGGAATGATATCACTATCAAAAGATTTGGAAGACATTTGTTTACT 4975
Db 1086 CAAGCCCTAGGAGATGATGACATTAAGTATTTCTCTTGAAGAAATTCCTC 1027
Qy 4976 CAAGTTTACCAATGATGTTTATTTTCACTTAATTAATGAAGAGGATCAGTAATAT 5035
Db 1026 AGAGGTAAAGTGATACCACTTA-TTCATTAAGAGAGCATATATATTTTGTGGT 968
Qy 5036 TTAGGTTATATGTTGATGATATCTTATGTTGGAAGTTCACAAAAGTATTTAGTAAT 5095
Db 967 TCA--AATATATGTTGATGATATTTTGGATCCATATGATTTATTTGTGCAAGAG 910
Qy 5096 TTTGTGATCAATTGAGATCATTTTGAAGTTAAAGTTTGTGTAATATCAAAATTAAT 5155
Db 909 TTTTCCCTTGATATGCAAGATGAAATTTGAATGTCAATGATGGAGAACTPAAAGTCTTT 850
Qy 5156 CTGGTATTTGAATTTCTPAAACCGAATCTGTTATTTTATCTCAAGAAATTTCTC 5215
Db 849 CTGGATTTACAAATACCAACTCAAGAGGTATTTATTCATCAATCAATCAATCTGC 790
Qy 5216 AAGAAATTAATTAAGATTTCAAACTAGATGACTCATATGAGAAAAACATACCTGATTT 5275
Db 789 AAGGATTTGATCAAAAGATTTGGGATGATGTCACAAACACATGCTACACCGATAGC 720
Qy 5276 CCGAATGACAAATATGAAAGTTTGA 5302
Db 729 ACTAATTTGTTACTTATGATTAAGATGAA 703

```

# RESULT 9

US-10-312-841-1/c  
 ; Sequence 1, Application US/10312841  
 ; Publication No. US20030186277A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Epigenomics AG  
 ; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC  
 ; FILE REFERENCE: E01/1208/WO  
 ; CURRENT APPLICATION NUMBER: US/10/312,841  
 ; CURRENT FILING DATE: 2002-12-30  
 ; NUMBER OF SEQ ID NOS: 2

; SEQ ID NO 1  
 ; LENGTH: 3673778  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 ; NAME/KEY: unsure  
 ; LOCATION: (3294164)  
 ; US-10-312-841-1

Query Match 1.1%; Score 68.4; DB 14; Length 3673778;  
 Best Local Similarity 59.1%; Pred. No. 0.12;  
 Matches 117; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```

Qy 1181 AACAAATATGAACTTAATCTTATTTAGTTATGTTTACACAGACGAGAAACCAAGAAAA 1240
Db 2552950 AACAACTCTAACCTTTCTTTCTCTAATTAATTAATTAATTAATTAATTAATTA 2552891
Qy 1241 CCAGAGGAACTCACTGGAACAAATCTCAAAAGAAAGAACTGAAATCAAGAAAGAAAT 1300
Db 2552890 AAATATCAAAACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2552831
Qy 1301 AAGAAACATCCAAATCATATACATTAAGTTAAAGTAAAAAGAAAAAGAAAAAACT 1360
Db 2552830 AAAAAAAAAAAAAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2552771
Qy 1361 TCATCGAATGAAAAACA 1378
Db 2552770 AAAAAAAAAAAAAAAAAA 2552753

```

## RESULT 10

US-10-311-455-1610/c  
 ; Sequence 1610, Application US/10311455  
 ; Publication No. US20030143606A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: OLEK, Alexander  
 ; APPLICANT: PIPEBEROCK, Christian  
 ; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Methylation Status of Cytosine Residues  
 ; FILE REFERENCE: 5013.1014  
 ; CURRENT APPLICATION NUMBER: US/10/311,455  
 ; CURRENT FILING DATE: 2002-12-16  
 ; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
 ; PRIOR FILING DATE: 2001-07-02  
 ; PRIOR APPLICATION NUMBER: DE 10032529.7  
 ; PRIOR FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: DE 10043826.1  
 ; PRIOR FILING DATE: 2000-09-01  
 ; NUMBER OF SEQ ID NOS: 2424  
 ; SEQ ID NO 1610  
 ; LENGTH: 7306  
 ; TYPE: DNA  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 ; US-10-311-455-1610

Query Match 1.1%; Score 68.2; DB 14; Length 7306;  
 Best Local Similarity 50.1%; Pred. No. 0.0064;  
 Matches 169; Conservative 0; Mismatches 168; Indels 0; Gaps 0;

```

Qy 1170 ACATTCAGACAAATATGATGATTAATCTTATTTAGTTTCCAGACACAGAGAAAC 1229
Db 6418 ACACCTTAATTAACCAACCACTTTATCTTAATTAATTAATTAATTAATTAATTA 6359
Qy 1230 CAAAAGAAAAACAGAGAGAACTCATGGAACATCTCAAAAGAAAGAACTGAAATCA 1289
Db 6358 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6299
Qy 1290 GAAAGAAATTAAGAAATCAATCAAAATCATGATTAAGATTAAGATTAAGATTAAG 1349

```

```
Db 6298 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 6239
Qy 1350 AAGAAAAAAGCTCACTGATGATGAAAAAGAGTGTCTCTTATTAATGTAATGATA 1409
Db 6238 AATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6179
Qy 1410 TACATTAATGACGAAAAACCGTTTCCAGTAGAAAAATTCATCTCTTATGCTT 1469
Db 6178 AATTATTAACCTCTCTTAATAAATAATCCAAATAAATAATTTAATCAATTAAT 6119
Qy 1470 TGAAGCTATGATTTTAAGTTTAAAGTTTAACTTAACAA 1506
Db 6118 TTATTAATAATTTAAAAAATAATATATCTTACTTA 6082
```

```
RESULT 11
US-10-312-841-2/c
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/NO
; CURRENT APPLICATION NUMBER: US/10/312,841
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2
```

```
Query Match 1.0%; Score 65.8; DB 14; Length 3673778;
Best Local Similarity 52.3%; Pred. No. 0.39;
Matches 145; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
```

```
Qy 1170 ACATTCGACGACCAATATGAACTTAATCTTATGTTTACCAAGCAGCAGAAAC 1229
Db 3287077 ACAACCTAAATATATAACCAACCTTATCTCAAAAAAAAAAAAAAAAAA 3287018
Qy 1230 CAAGAAGAAACCAAGAGAACTGACGAAACAATCCAAAGAAAGAACTGAAATCA 1289
Db 3287017 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3286958
Qy 1230 GAAAGAGAAATAGAAACATCCAAATCAGATAACGATAAAGTGAAAAAGAAAAA 1349
Db 3286957 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 3286898
Qy 1350 AAGAAAAAAGCTCACTGATGATGAAAAAGAGTGTCTCTTATTAATGTAATG 1409
Db 3286897 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATTTTCAATCAATTAATA 3286838
Qy 1410 TACATTAATGACGAAAAACCGTTTCCAGTAGAAAA 1446
Db 3286837 CATATTAATCAAAATAAACCATTTATAAAAAAAAAA 3286801
```

```
RESULT 12
US-10-424-599-108640
; Sequence 108640, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
```

```
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 108640
; LENGTH: 7628
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (7628)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_69115C.1
US-10-424-599-108640
```

```
Query Match 1.0%; Score 65.6; DB 12; Length 7628;
Best Local Similarity 45.1%; Pred. No. 0.021;
Matches 371; Conservative 0; Mismatches 439; Indels 12; Gaps 3;
```

```
Qy 4484 AATGATCCGAGATGATTAAGTCATGAATGCTGAATCTAGAGAAATTTGATCAAAAGT 4543
Db 4263 ACTGATGAGTTCTGGATGATGCTATGCAAAAAAGAAATGAGCAATCAAAAGATGA 4322
Qy 4544 GTTTACGAAGAGTTCCAAATTCACCGGTGTAAGCCATATATCTATGAGGTGGATACAT 4603
Db 4323 GTCTGGAGCTAGTTCTTATGACCCAGGAACTATATGATGATGACCAAGTGTCTTC 4382
Qy 4604 ACTGAGAAATGATTTCTCAAAAGGTGTGTCG---GAAATCAAGTTGTGTGCAT 4660
Db 4383 AAGAACAAACCAATGAAGAGGTGTATACAGAAACAGCCAGACTGTGTGCTCA 4442
Qy 4661 GCGAACAGCAAAAGAAAAATTTGATTTGACCCCTTATAGCTTAGTTCACCTGTATA 4720
Db 4443 GGCTACACTGATTAAGGGGAGCTTTGA-----TGAAACTTTCGCCCCGTGTGCT 4496
Qy 4721 GATCTTGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 4780
Db 4497 AGACTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4556
Qy 4781 CATTTAGAGTGCAGTGGCGATCTTAATGCTCTATTTACATTCAAATCCATTTAT 4840
Db 4557 CAGATGATGTGAAGCGCGCTTTCGAATGATACCTGAATGAAGAACTGATGAGAG 4616
Qy 4841 GTCTTCCTCTTAATCACTACCTTTGAAGAAACCAATGTTGTTGTTATGAAGCTTC 4900
Db 4617 CAGCAAGAGATTTCAATCACTCACTCACTCACTCACTCACTCACTCACTCACTCA 4676
Qy 4901 GTCTATGAGTTAAACAGTGGGTTGATGATGATGATGATGATGATGATGATGATGAT 4960
Db 4677 CTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4736
Qy 4961 GACATGTTTAACTAAGTTTAACTAAGTTTAACTAAGTTTAACTAAGTTTAACTAAG 5020
Db 4737 CAGCAAGGATTAAGAGGAGGAGGATTAAGCAACTCTTTTGTCAACAAAGTGTGA 4796
Qy 5021 GATCAGTAATATTTAGTTTAACTGATGATGATGATGATGATGATGATGATGATGAT 5080
Db 4797 AACTGATGATA---GCAAGATATATGTTGATGATGATGATGATGATGATGATGAT 4853
Qy 5081 AAGATTAATGATTAATTTTGTGATCAATGAGAGATCAATTTGAAGTTAAAGTTTGT 5140
Db 4854 GAGATGCTTGAATTTTGTCAACAGATGATGATGATGATGATGATGATGATGATGAT 4913
Qy 5141 GAAATATCAATTAATCTTGTGATGATGATGATGATGATGATGATGATGATGATGAT 5200
Db 4914 GAGCTACTTATTTCTGAGACTTCAAGAGAGAGATGAGAGATGAGAGATGAGAGATGAG 4973
Qy 5201 CAAAGAAAAATTTCTCAAGAAATTTACTTAAGATTTCAAACTGATGATGATGATGAG 5260
Db 4974 CAAAGAGGATGACAAAGAACTGTGCAAGAGTTGGATGAGAGATGAGAGATGAGAGAT 5033
```



Mon Apr 12 10:25:52 2004

us-09-430-590e-3.rnpb

Page 9

[illegible]

Search completed: April 12, 2004, 01:18:39  
Job time : 1460 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 11, 2004, 13:02:12 ; Search time 9825 Seconds  
(without alignments)  
19531.234 Million cell updates/sec

Title: US-09-430-590E-3

Perfect score: 6426  
Sequence: 1 tgcgtgttcgtgcacattc.....agaagtattatccatca 6426

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estdb:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estum:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vit1:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	174.6	2.7	1017	29	CNS07BJD AL437951 T7 end of
2	142.4	2.2	905	29	CNS07JAX AL436423 T3 end of
3	127.2	2.0	838	29	CNS07JAX AL436955 T7 end of
4	103.6	1.6	781	12	BI920632 EST540579

5	97.8	1.5	569	9	AI813079
6	96.2	1.5	1035	29	CNS07JAX
7	93.6	1.5	832	14	CB970396
8	92.8	1.4	766	29	CG964290
9	91.4	1.4	866	29	CG937798
10	91.4	1.4	867	29	CG973377
11	91.4	1.4	869	29	CG961676
12	89.6	1.4	940	29	CG928381
13	89.4	1.4	803	29	CG821291
14	89.2	1.4	926	29	CG943796
15	88.6	1.4	490	29	AG264218
16	88.4	1.4	714	28	CG917427
17	87.6	1.4	785	28	BH664738
18	87.6	1.4	901	28	CG944168
19	87	1.4	780	28	BH484021
20	86.8	1.4	1200	13	BX437758
21	86.6	1.3	1098	13	BX377526
22	86.4	1.3	765	28	BH507520
23	86.2	1.3	688	13	BQ510267
24	86.2	1.3	721	28	BH723836
25	86.2	1.3	784	28	BH445102
26	86	1.3	827	29	CG975352
27	86	1.3	846	29	CG935096
28	86	1.3	852	29	CG932551
29	86	1.3	857	29	CG923622
30	86	1.3	858	29	CG932091
31	86	1.3	875	29	CG971290
32	86	1.3	890	29	CG934171
33	86	1.3	923	29	CG941085
34	85.8	1.3	829	29	BX173672
35	85.4	1.3	712	29	CG967635
36	85.2	1.3	844	29	BX139887
37	85.2	1.3	914	29	CG956730
38	85	1.3	827	29	CG961943
39	85	1.3	905	28	AZ550256
40	84.8	1.3	638	29	AG251630
41	84.6	1.3	1130	28	BZ433063
42	84.4	1.3	808	28	BZ505487
43	84.2	1.3	914	29	CG971731
44	84	1.3	700	28	BH728433
45	84	1.3	945	29	CG936465

## ALIGNMENTS

RESULT 1	CNS07BJD	1017 bp	DNA	linear	GSS 08-JUL-2001
LOCUS	T7 end of clone BC0A012E08 of library BC0A from strain CBS 767 of				
DEFINITION	Debaromyces hansenii, genomic survey sequence.				
ACCESSION	AL437951.1	GI:12221364			
VERSION	AL437951.1				
KEYWORDS	GSS.				
SOURCE	Debaromyces hansenii (anamorph: Candida famata)				
ORGANISM	Debaromyces hansenii				
REFERENCE	Saccharomycetales; Saccharomycetaceae; Debaromyces.				
AUTHORS	1 (bases 1 to 1017) Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., deMontigny,J., Dujon,B., Durans,P., Lepingle,A., Lorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Poirier,S., Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Jouvenel,M., Wincker,P. and Weissenbach,J.				
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of				
JOURNAL	YEAST	18			
MEDLINE	20584711				
PUBMED	1152876				
REFERENCE	2 (bases 1 to 1017) Lepingle,A., Casaregola,S., Neuvéglise,C., Bon,E., Nguyen,H., Artiguenave,F., Wincker,P. and Galliardin,C.				





```

/misc_feature
/variety="hansenii"
/db_xref="taxon:4959"
/clone="BCOA003C05"
/clone_lib="BCOAA"
/note="end : T3"
<1..>905
/note="part of putative transposable element"
/evidence=not_experimental

```

## ORIGIN

```

Query Match      2.2%; Score 142.4; DB 29; Length 905;
Best Local Similarity 51.0%; Pred. No. 2.9e-15;
Matches 347; Conservative 5; Mismatches 327; Indels 1; Gaps 1;

4444 CAAGAGGATTTACATGCGGTGCTCAACAAATGACATATATATCCGGATGATAAA 4503
227 CACTCGGAAATTAATCTATATCGAAAGTCGATTTAAGATTAACAATGCGAGCA 286
4504 GTCAATGATGCTGAAGTGAAGAAATTTAGATCAAAAGTGTTCAGAGAGTTCCAT 4563
287 GTCAATACAGCGAATAGATCTTTTGAAGAAATATAGATATATACGGTTGGAAMA 346
4564 TCCACCGGTGTAAGCTTATATATGCGGTGCTATCTAGAGAAATTCATCTCT 4623
347 TCTTAATAATGTCGAACTTATTCACACTTGGGTTCATACACATAAAATTCGATCT 406
4624 CAAGAGTGTGTTGGAATCAGCTGTGTGTCATGCGCAACAGCAAAAGAAATTT 4683
407 CAAGAAGTTAGTAATAATCAGCTGTGTGTCAGAGGCTTTAGCAAAATTCGAATGA 466
4684 GGATTTATACCTTTTATGTTAGTTACCTGTTATATGATCTTTGCTATAGATTATT 4743
467 ACACATATATACCTCGAAGGTGTCATCTCCGTGATGATTATCCATAATTCGTTACT 526
4744 GACATATATAGTTGATGATTAAGAAATGACATTCATTCAGCTGCGGCGTA 4803
527 TACACCGATATAGT-TGAATATGAAATGCCCATACATCATTTATATCTCCGATA 585
4804 TCTAATGCTCTTATCTACTATTCATTCATTCATTCATTCCTCTCTTAATCAGTACC 4863
586 TTTACATGACATATGACATATGAGAAATCCATATTTGTAAACACACCTGATGAA 645
4864 TTGAGAGAAACCATGTTGTTATGAAACGTTGCTATGCGGTTAAACAGTGGG 4923
646 TATTGATTTCTGTAATGTTGCAATTAACAAATCTGTTATGGAATGAGACAGCAGG 705
4924 TTTGGAATGATGACATATCAAAAGATGATGGAAGACATTTGTTTACTCAAGTTT 4983
706 GTATATGATGATCAATGATATACAAAGTTCTTATGATCTTATTCGAACTGATAC 765
4984 ACAATATGATGTTTATTTTCAATGAAATATGAAGAGGATCAGTAATATTTAGTTT 5043
766 TGCCATTAAGCGAAATGTTTGTAATATTTTGTAAGAAATTAAGAGCTATCGTTGACAT 825
5044 ATATGATGATATATTTCTTATGTTGGAAGTTCAAAAGTTATATTTATTTTGA 5103
826 ATATGATGATATATTTTACTCTGCTAATATTAATTCATTTTCAAGATTTCAACT 885
5104 TCAATTGAGAGATATTTTG 5123
886 TGAACGCTGTAACATTTTCG 905

```

```

RESULT 3
LOCUS      CDS07ARP      838 bp      DNA      linear      GSS 08-JUL-2001
DEFINITION T7 end of clone BCOA006C12 of library BCOA from strain CBS 767 of
            Debaryomyces hansenii, genomic survey sequence.
ACCESSION  AL436955
VERSION     AL436955.1 GI:12220368
KEYWORDS   GSS
SOURCE      Debaryomyces hansenii (anamorph: Candida famata)
ORGANISM   Debaryomyces hansenii

```

## REFERENCE

Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Debaryomyces.

## AUTHORS

Souchet, J.-L., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S., de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, B., Malpertuy, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Portier, S., Sautin, M., Tekaia, F., Toffano-Nicolas, C., Wesolowski-Louvel, M., Wincker, P. and Weissenbach, J.

## TITLE

Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies

## JOURNAL

FEMS Lett. 487 (1), 3-12 (2000)

## MEDLINE

20584711

## PUBMED

1152876

## AUTHORS

Lepingle, A., Casaregola, S., Neuvéglise, C., Bon, E., Nguyen, H., Artiguenave, F., Wincker, P. and Galliard, C.

## TITLE

Genomic exploration of the hemiascomycetous yeasts: 14. Debaryomyces hansenii var. hansenii

## JOURNAL

FEMS Lett. 487 (1), 82-86 (2000)

## MEDLINE

20584724

## PUBMED

1152889

## REFERENCE

3 (bases 1 to 838)

## AUTHORS

Genoscope.

## TITLE

Direct Submission

## COMMENT

Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 Evry cedex - FRANCE (E-mail : segre@genoscope.cns.fr - Web : www.genoscope.cns.fr)

This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvarum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

## FEATURES

Location/Qualifiers

## source

1..838

## misc\_feature

/organism="Debaryomyces hansenii"

## ORIGIN

/mol\_type="genomic DNA"

## misc\_feature

/strain="CBS 767"

## misc\_feature

/variety="hansenii"

## misc\_feature

/db\_xref="taxon:4959"

## misc\_feature

/clone="BCOA006C12"

## misc\_feature

/clone\_lib="BCOAA"

## misc\_feature

/note="end : T7"

## misc\_feature

<1..>838

## misc\_feature

/note="part of putative transposable element"

## misc\_feature

/evidence=not\_experimental

## ORIGIN

Query Match 2.0%; Score 127.2; DB 29; Length 838;

## misc\_feature

Best Local Similarity 49.8%; Pred. No. 1.5e-12;

## misc\_feature

Matches 308; Conservative 12; Mismatches 298; Indels 1; Gaps 1;

```

4444 CAAGAGGATTTACATGCGGTGCTCAACAAATGACATATATATCCGGATGATAAA 4503
218 CACTCGGAAATTAATCTATATCGAAAGTCGATTTAAGATTAACAATGCGAGCA 277
4504 GTCAATGATGCTGAAGTGAAGAAATTTAGATCAAAAGTGTTCAGAGAGTTCCAT 4563
278 GTCAATACAGCGAATTAATGATCTTTTGAAGAAATTAAGATTAATAGGTTGTAAT 337
4564 TCCACCGGTGTAAGCTTATATCTATGCGGTGCTATCTAGAGAAATTCATCTCT 4623
338 TCTTAATAATGTCGAACTTATTCACACTTGGGTTCATACACATTAATTAATTAAGATCT 397
4624 CAAGAGTGTGTTGGAATCAGCTGTGTGTCATGCGCAACAGCAAAAGAAATTT 4683
398 CAAGAAGTTAGTAATAATCAGCTGTGTGTCATGCGCAATTCGCAAAATTCGAATGA 457

```

QY	4664	GGATATATACACCTTTTACTGTAGTTCACCGTGTATAGTCTTGATCTGGACTATTAAGATTATTT	4743
Db	458	ACACTATATATACCTCGAAGGTGTATCTCTCGTGAATTGACCTATTCATAATTCGTTTACT	517
QY	4744	GACAAATATAAGGTTGTGTAATTAGGAATGACAATTCACATTAACTTTAAGCGTGGAGTGCTA	4803
Db	518	TACAGCGATAGAGT-TGAATATGACTGGCCGATACATCATCTTGATATATCTCCGGCAATA	576
QY	4804	TCTAAATCCCTCTATTACTCATTTCCAAATCCCAATTTATCTCTTCTCCTTAATACAGTAC	4865
Db	577	TTTCCCTGCCCATATCGACTACAGCTACAGTCATACAGCAATATATTTGACCCCAACCTGGCTMGAC	636
QY	4864	TTTGAGAGAAAACCACTGTGGTGTATTGAAAGCTGTCTGCTATAGGGTTAAACAAGTCCGG	4923
Db	637	TATGATATCTCGTAAATATTTGGNAATTCMCCCCACCTGTATATSCATGACACAAGACAG	696
QY	4924	TTTGGAAATGATACACACTATTCAAAAGAGTATGGAAACAATGGTATTACTCAAGTTT	4983
Db	697	GTATATGTGATCAATCATGTATATTAACAGAGTCTTATGGCTCTTAATTTCGAAMCTGATAM	756
QY	4984	ACACATATATGTTTATTTCACATTTGAATATGAAGAGGATCAGTAATATATTAGGTTT	5043
Db	757	TGCSATTTGCGGAAGTATTTGTAATATTTTGTGTAATAATGAAGAGCTCATCGTTCACCT	816
QY	5044	ATATGTGATGATTTCTT	5062
Db	817	ATATGTGATGATATGTTT	835

RESULT 4	751 bp	mRNA	linear	EST 10-MAR-2003
B1920632				
B1920632				
LOCUS				
DEFINITION	B1920632	potato microtublers, in vitro-grown	Solanum tuberosum	CDN
	EST140573	clone CSTE26G17 5' end, mRNA sequence.		

ORCHARDTUM	
SOURCE	Solanum tuberosum (potato)
KEYWORDS	EST.
VERSION	BI920632.1 GI:16216672
ACCESSION	BI920632

**ORGANISM**  
Solanum elaeagnifolium  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
SpERMATOPHYTES; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamiales; Solanales; Solanaceae; Solanum.  
**REFERENCE**  
1 (bases 1 to 781)  
**AUTHORS**  
van der Hoeven, R., Bezzerides, J., Bachem, C., Visser, R.,  
Thaler, D., Thaler, T., Chikwino, A.

TITLE  
 Generation of ESs from in vitro grown microtubers (2011b)  
 JOURNAL  
 Unpublished (2001)  
 COMMENT  
 Contact: Robin Buehl

The Institute for Genomic Research  
9717 Medical Center Dr, Rockville, MD 20850, USA  
Email: [porcato-array@igr.org](mailto:porcato-array@igr.org)  
This clone can be obtained from the University of Arizona Genomics  
Institute. Orders can be made through URL:  
<http://genome.arizona.edu/orders/>  
Seq primer: T3.

FEATURES	Location/Qualifiers
source	1. .781

```

/organism="Solanastrum tuberosum"
/mol_type="cRNA"
/cultivar="Bintje"
/db_xref="taxon:4113"
/clone="c3152617"
/tissue_type="axillary buds of stem explants; growing
sink-tubers"
/dev_stage="7, 8 and 10 days"
/lab_host="SOLR"
/clone_lib="potato microtubers, in vitro-grown"
/notes="Vector: pInscript SK(-); Site_1: EcoRI; Site_2:
XhoI; Tissue supplied by Christian Bateman and Richard
Visser (Department of Plant Breeding, Wageningen
University, The Netherlands); sequencing by The Institute
for Genomic Research. The cDNA libraries will attempt to

```

capture the induction and initiation/initial growth of the tuber in an in vitro system as described in Bohem et al. (Plant Journal, 1996). Small microtubers develop from axillary buds attached to stem explants when placed on a high sucrose medium (10%). Visible morphological changes occur synchronously at day five in the axillary buds. The first library, cSTa (1-20) consists of axillary buds harvested on days 1-3. This targets those genes involved in induction of the microtubers. The following libraries, cSTb (21-40) and cSTc (41-60), capture genes involved in tuber initiation and outgrowth. This library is noted as p3 in Tanksley lab notebooks."

Query Match 1.6%; Score 103.6; DB 12; Length 781;

Best Local Similarity 48.0%; Prec. NO. 2.3e-08;  
Matches 360; Conservative 0; Mismatches 384; Indels 6; Gaps 2;

4486 TGATCCGGATGATAAGTCATGAATGCTGAAC TAGA GAATTTAGATCAAAAGATGT 4545

D6 15 TCATTGTGAGGAGGAGGCTATGTTGATGAGATGTCTGCTTTACATAAGAGTGTTAC 74

4546 TTACGAGAGCTTCCAATCCACCGGTGTGAAGCCTATATCTATGGTTGGGTACACTAC 4603

Db 75 TTGGGAGCTTGTCTCCCTTCCCTGCAGGAAATCTACTGTGGTTGTTCGTTGGGTTTAGC 134

[illegible]

Db 135 TATCAAATTGGTCAGACGGTCAGGTTGATCGACTTAAAGGCTCCGCTTGT---TCCCAA 191

4666 CAGACAAAGGAAAMTGGATTATGACCCCTTTAGTGGTGGTCCCTGTTATGAGCACT

4736 TCGAGTATTAAGATTATGACATAATAGCTTGGAAATTAGGAATGACAAATCAACATTT 4785

252 TGCATGTGCTCTTTCTCTAATCTAGCTGCCGTCGTCATTGGCCTCTTCATCAGTT 311

4786 AGACGTCAGTCGGCGTATCTAAATGCGCTCTATTACTCATTCAAATCCATTATATGCTT 4845

Db 312 GGACATTAGGAATGCTTTTCTGCATGGTTATCTTGAGGAGAAGTCTATATCGAGCAACC 371

4846 TCCTCTAATCAGTACTTTGAAG--AAAAACATTGTTGTTAATTGAACGTTCTGT 4902

Db 372 ACCGTGTTTGTTCCTAGGGGAGTCTAGTACCTTGATGTCGATTGCCAGGTACT 431

4903 CTATGGGTTAAACAGTCGGGTTTGGAAATGCTATCACTATCAAAAGACTATGGAGA 4962

Db 432 CTATGGCTCTAAAACAGTCTCCTCGAGCTTGCTTGGGAATTCACACAGTATTTCAGGA 491

QY 4963 CATTGGTTTACTCAAGTTTACACATGATGTTATTTCACATTCGATATGAAGAGGG 50222

Db 492 GTTGGATGACTCGTAGTGAGCTGATCACTGTGTATTATCGGCATTTCACCAAG 551

5023 ATCAGTAATAATTTAGGTTTAAATGTTGATGATATCTTAATGTTGGAAGTTCACAAA 5082

Db 552 TCATGTAICTATTGGTGTGTTATGTTGATGATATGTTATCACCAGTAATGATCAAGA 611

5083 ACTATTCGATATTTTTCGATCATTCGAGAGATTTTTCGAGTTAAAGTGTTCGCGA 5142

612 TGGTATCACCGATTTGAGCACAATCCTTTAAAGCACCTCCAGACCTTAAGACCTGGCAG 874

5143 AATATCAATATATCTTGGATATGATATGGTAAATCTGATAGATACAGGATATGTTATTTTCA 734

5203 AGGAAAAATTTCTCAAGAAATTACTTAAGGA 5232

Db 732 ACGCAAGTATGCCCTTAGACATTCGTGAGGA 761

RESULT 5  
A1813079

LOCUS A1813079 569 bp mRNA linear EST 08-JUL-1999  
 DEFINITION 20E11 Pine Lambda Zap Xylem library Pinus taeda cDNA, mRNA  
 ACCESSION A1813079  
 VERSION A1813079.1 GI:5424175  
 KEYWORDS EST.  
 SOURCE Pinus taeda (loblolly pine)  
 ORGANISM Pinus taeda (loblolly pine)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 1 (bases 1 to 569)  
 REFERENCE Whetten, R.W., Kianlay, C.S., Rezzel, E. and Sederoff, R.R.  
 TITLE The Pine Gene Discovery Project  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Rose Whetten  
 Forest Biotechnology Group  
 North Carolina State University  
 Dept. of Forestry, NC State University, 6113 Jordan Hall,  
 Raleigh, NC, 27695-8008  
 Tel: 919-515-7800  
 Fax: 919-515-7801  
 Email: rosswhetten@ncsu.edu  
 Seq primer: T3  
 Location/Qualifiers  
 1..569  
 /organism="Pinus taeda"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:3352"  
 /tissue\_type="differentiating xylem"  
 /clone\_lib="Pine Lambda Zap Xylem library"  
 /note="Vector: Lambda Zap; Site 1: EcorI; Site 2: XhoI;  
 Differentiating xylem was collected from the main stem of  
 a 35-year old loblolly pine tree harvested during the  
 growing season. RNA isolation and library preparation  
 followed the methods of Allona et al., FNAS 95:9693-8,  
 1998"

ORIGIN  
 Query Match 1.5%; Score 97.8; DB 9; Length 569;  
 Best Local Similarity 48.5%; Pred. No. 2.6e-07;  
 Matches 270; Conservative 0; Mismatches 287; Indels 0; Gaps 0;

QY 4730 ACTTAAAGATTATGACATATATAGTTGATTAAGAAATGACATTCACATTAGAC 4789  
 Db 13 ACAATTGATTACGATTCCTTAGCCACAAACATATTGGAACATACATCAATTGAT 72  
 QY 4790 GTCCAGTGGCGTACTTAATGCTTATCTACATTCATCAATTCATTTATGCTTCC 4849  
 Db 73 GTCAAAATGCTTTCTTAATGATGATTTAAGGAAGAATTTATTTGTTCAACTGAA 132  
 QY 4850 CCTAATCAGTACCTTTGAAGAAAACCATTTGTTTATGAACGTTCTGTATGG 4909  
 Db 133 GGGTTTGCAACAGGACAGGACAGCATCTGGTTGACGGTTAAGAAAGCATTTGAT 192  
 QY 4910 TTAAGACAGTGGGTTTGATGATGATCTTCAACATTCAGATTTGGAAGACATTGT 4969  
 Db 193 TTGAACAGGACCAAGGTCATGATGTAAGATGATGATCTTTCTTCGAATGGA 252  
 QY 4970 TTTAAGCAAGTTTACCAATGATGTTTATTTTCACTTGAATTAAGAGGATCAGTA 5029  
 Db 253 TTTGTGAAGACAAAATGATCTTAAGATGATTAAGAAAGTGAAGAAATGTT 312  
 QY 5030 ATATATTTAGTTTATATGTTGATGATTTCTTATGTTGGAAGTTCACAAAAGTTAT 5089  
 Db 313 GCTTTAATATCTTTGATGATGATGATCTTAATTAACAGGTGATGCTCAAGTTAAT 372  
 QY 5090 GATAATTTTGATGATGATGATGATGATTTGATGATGATGATGATGATGATGATGAT 5149  
 Db 373 GAAGAAATTAAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAG 432  
 QY 5150 AATTATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5209  
 Db 433 TACTGTTAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 492

QY 5210 TTTTCAGAAATTAAGATTTCAGATGATGATGATGATGATGATGATGATGATGATGAT 5269  
 Db 493 TATACAAAGAGATTCCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 552

QY 5270 TGGATTCGGAATGACA 5286  
 Db 553 TTGAACAGAAATGACAA 569

RESULT 6  
 CDS7AOP/c  
 LOCUS T7 end of clone BC0A005F10 of library BC0A from strain CBS 767 of  
 DEFINITION Debaryomyces hansenii, genomic survey sequence.  
 ACCESSION AL436847  
 VERSION AL436847.1 GI:12220260  
 KEYWORDS GSS.  
 SOURCE Debaryomyces hansenii (anamorph: Candida famata)  
 ORGANISM Debaryomyces hansenii  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 1 (bases 1 to 1035)  
 REFERENCE Soucier, J., Aigle, M., Artiguenave, F., Blandin, G.,  
 Bolotin-Fukuhara, M., Bon, E., Brotier, P., Casaregola, S.,  
 de-Montigny, J., Dujon, B., Durieux, P., Lepingle, A., Llorente, B.,  
 Malpertuy, A., Neuvéglise, C., Olier-Kalogeropoulos, O., Potier, S.,  
 Saurin, M., Tekala, F., Toffano-Nioche, C., Weislowski-Douvet, M.,  
 Wincker, P. and Weissenbach, J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEMS Lett. 487 (1), 3-12 (2000)  
 MEDLINE 20584711  
 PUBMED 11152876  
 REFERENCE 2 (bases 1 to 1035)  
 AUTHORS Lepingle, A., Casaregola, S., Neuvéglise, C., Bon, E., Nguyen, H.,  
 Artiguenave, F., Wincker, P. and Galliard, C.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 14.  
 Debaryomyces hansenii var. hansenii  
 JOURNAL FEMS Lett. 487 (1), 82-86 (2000)  
 MEDLINE 20584724  
 PUBMED 11152889  
 REFERENCE 3 (bases 1 to 1035)  
 AUTHORS Genoscope.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Séquençage,  
 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex - FRANCE (E-mail :  
 seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)  
 COMMENT This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces  
 exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,  
 Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces  
 lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia  
 angusta, Debaryomyces hansenii var. hansenii, Pichia scottiiophila,  
 Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.  
 Location/Qualifiers  
 1..1035  
 /organism="Debaryomyces hansenii"  
 /mol\_type="Genomic DNA"  
 /strain="CBS 767"  
 /variety="hansenii"  
 /db\_xref="taxon:4959"  
 /clone="BC0A005F10"  
 /clone\_lib="BC0A"  
 /note="Tend : 77"  
 <1..>1035  
 /note="Part of putative transposable element"  
 /evidence="not\_experimental"

misc\_feature  
 ORIGIN

vines were located at the University of California, Davis Experimental Vineyard. cDNAs were made by oligo-dT priming and directionally cloned. 5' and 3' adaptors were used in cloning as follows:

5'-AAGCAGTGGTATCAACAGAGTGGCCATTACGCGCGG-3' and 5'-ATTCTAGAGCGCGGAGCGCGGACATG-dT(30)NN-3'. Library was constructed using the Clontech Creator (SMART) kit and size-selected to contain the 0.5-3 kb size fraction. "

Query Match	1.5%	Score 93.6;	DB 14;	Length 832;
Best Local Similarity	47.4%;	Pred. No. 1.4e-06;		
Matches 279;	Conservative 0;	Mismatches 309;	Indels 0;	Gaps 0;

Qy	4708	TTCCACCTGTTATTAAGCTCTTGACCTATTAAGATTATGACAAATATAGGTTGGAATTAGG	4767
Db	602	TGCTCCTCTTGCAAAACCTGAACACTATCAGAGATCCCTTCCTCATTTGGCTGCATCAAGA	543
Qy	4768	AATGACAATTGAACATTTAAGTCGACGTCGGGGATATCTAAATGCTCTATTAATCTATTC	4827
Db	542	TTGGTGCTTGGACAACCTGACATPAAAAAATGGGTTCTTAAATGGGGACCTGGAGAGGA	483
Qy	4828	AAATCATTAATTATCTCTTTCCTCCATAACAGACTTTGGAAGAAAAACCTTTTGTT	4887
Db	482	AATCTACATGGAATTAACCCCTCGTTTGCAGAAAGATATGGCAATCAGCTTTCGA	423
Qy	4888	ATTGAACGTTTCTGTCTATAGGTTAAACAAGTCGGGTTTGGATAGTATCACTATCA	4947

QY 4948 AAGAGTATTGGAAACATGGTGTTCCTACGTTTACACATGATGGTATTTCACAT 5007

Db 362 AAAACACATCTCTGAAGTGGGTACAAACAAGTCAGGCTCATCTCATCTCATTTTGCAA 303

QY 5008 TAAATATACAGGATCAGTAATTAATTAGTATTATCTGATGATATCTTAAGT 5067

Db 302 GAGAGCTATGCCGGAAATTGGCCCATTTGATTGATCTATGTCAGTATATTTCATC 243

QY 5068 TCGAAGTTCACAAAAGTTATTGATATTGATTTTGATCAATTGGAGACATATTTGAGT 5127

Db 242 TGGGAGATATATGGGGAGGTACAGAAATTGAAATAATTTGTCAAGAGGTTTGAGT 183

Qy 5128 TAAAGTGTTCGTGAAATATCAAAATATCTTCGTCTTGAATTTGCTAATAAACCAATCTG 5187

Db 182 TAAACCCCTTGAAATTTGAAATATTTCTCTTGATATGAAAGTCTTGATCAAGAAAGG 123

Qy 5188 TTATATTTTATCTCAAGAAAAATTTCTCAAGAAATTTCTTAGGATTTCAACTAGATGA 5247

122 AATGATGATCTCTCAAGAAATATCATCTGATCTTTTAGAGAGCCCGTATGCTTG 63

Qy	5248	CTCATATGGGAAAAACATACCCGATTCGCAATGACCAATATATATA	5295
Db	62	ATGCMAACCAATTGATCTCTATGATGACGAAAAAAAAAAAAAAAA	15

RESULT 8  
CG364290 766 bp DNA linear GSS 15-DEC-2003

ACCESSION	Survey sequence.
VERSION	CG964290
KEYWORDS	CG964290.1 GI:39888829
SOURCE	GSS:
ORGANISM	Medicago truncatula (barrel medic)
REFERENCE	Medicago truncatula
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
JOURNAL	Rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliace;
	Medicago.
	1 (bases 1 to 766)
	Town,C.D., Sherry,J., Koo,H. and Feldhym,T.F.
	Sequencing of BAC ends from Medicago truncatula
	Unpublished (2003)

## COMMENT

Other GSSs: MBECD18TRC  
 TIGR  
 Contact: Chris Town  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdlowm@tigr.org  
 Seg primer: CAGGAAACGCTAGACG  
 Class: BAC ends

## FEATURES

source

Location/Qualifiers

1..766  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="35D12"  
 /clone\_lib="mth2"  
 /note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Cook, D.R. and Kim, D.J, unpublished"

## ORIGIN

Query Match 1.4%; Score 92.8; DB 29; Length 766;  
 Best Local Similarity 47.0%; Pred. No. 1.9e-06;  
 Matches 286; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

4725 TTGTGACTATTAAGATTATTAAGCAATAATAGGTTGTGAATTAGCAATTCACATT 4784  
 73 TGGATACATTCGAATTCCTGTAGCGTTGGCAGTACAAATAAGTGAAATTGGCATT 132  
 4785 TAGACGTGAGTCGGCGGATTAATGCTTATTACTCATTCACATTTATGTCT 4844  
 133 TAGATTTCAAATCGACATCTTTAAATGGAATCTGTAGAGAAATTATGTGCTCAC 192  
 4845 TTCCCTCTAATCGATCCTTTGAAGAAAACCATTTGTGGTTATTTGAACGTTCTCT 4904  
 193 CTGCTGCTTTTGTGTAAGAGGAGAGACAGGTATTAAGCTTCATTAAGCTTTGT 252  
 4905 ATGGGTTAAACAGTCGGGTTGGAATGCTATCACTATCAAAAGAGATTTGGACACA 4964  
 253 ATGGGCTGAAGACGGCCCTAGAGCTGTGCAACAAATTAATATCACATTTCTTAATC 312  
 4965 TTGGTTTACTCAAGTTTACACATGATGGTTTATTTACATGGAATATGAAGGGAT 5024  
 313 AAGGATTAAGAGAGTAAATGATGCACTTTATATGAAGAGCTTTGTGATGTG 372  
 5025 CAGTAATATATTAGTTTATATGTATGTATGATTTATTTATGTTGAAGTTCACAAAAG 5084  
 373 GTTCCTTATATGCTCTTTGATGTGATGACCTGTATGACATGAACAAATATATCAACAG 432  
 5085 TTATTGATATTTTGTGATCAATTTGAGAGTCATTTTGAAGTTAAAGTGTGTGTA 5144  
 433 AAGTACATCAACTTATGAGAGATGAAAACAGTTGAGATGCTAGCTTAGGGGAAA 492  
 5145 TATCAAAATATCTTGTATGATTTGTAACACCAATCGTTATTTATCTCAAG 5204  
 493 TGAACATATTTCTTGCTTGAAGATGATCAATCTGAGAGTGAATTTTTTGAATCAAG 552  
 5205 AAAAATTTCTCAGAAATTTACTTAAGATTTCAAACTAGATGATCTCATATGGGAAAACA 5264  
 553 AGAAGAAATGCTCAGAAATTTTGAAGAGTTTAAATGGAAGCTGCAAAATCTGCTCAA 612  
 5265 TACCTGGATTCGGAATGACAAATATGAAAGGTGCAATATCTGGAAGAGCTTAATC 5324  
 613 CTCCTTTGTGTGAATTTGAATCTTCAAGGAAGATGAAGCTGATTAATATGATGCTT 672  
 5325 CAGAGAAAT 5332  
 673 CTCATTAT 680

RESULT 9  
 CG937798 866 bp DNA linear GSS 12-DEC-2003  
 LOCUS CG937798

## DEFINITION

MBECA46TF mth2 Medicago truncatula genomic clone 24G19, genomic survey sequence.

ACCESSION  
 VERSION  
 CG937798.1 GI:39809821  
 GSS:  
 KEYWORDS  
 SOURCE  
 ORGANISM

Medicago truncatula (barrel medic)  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.

REFERENCE  
 1 (bases 1 to 866)  
 TOWN, C.D., SHELLEY, J., KOO, H. and FELDBLYUM, T.F.  
 TITLE  
 JOURNAL  
 Unpublished (2003)  
 COMMENT  
 Other GSSs: MBECA46TR

Contact: Chris Town

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdlowm@tigr.org

Seg primer: TGTAAACGACGGCCAGT  
 Class: BAC ends.

## FEATURES

source

Location/Qualifiers

1..866  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="24G19"  
 /clone\_lib="mth2"  
 /note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Cook, D.R. and Kim, D.J, unpublished"

## ORIGIN

Query Match 1.4%; Score 91.4; DB 29; Length 866;  
 Best Local Similarity 47.5%; Pred. No. 3.4e-06;  
 Matches 272; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

4725 TTGTGACTATTAAGATTATTAAGCAATAATAGGTTGTGAATTAGCAATTCACATT 4784  
 73 TGGATACATTCGAATTCCTGTAGCGTTGGCAGTACAAATAAGTGAAATTGGCATT 132  
 4785 TAGACGTGAGTCGGCGGATTAATGCTTATTACTCATTCACATTTATGTCT 4844  
 133 TAGATTTCAAATCGACATCTTTAAATGGAATCTGTAGAGAAATTATGTGCTCAC 192  
 4845 TTCCCTCTAATCGATCCTTTGAAGAAAACCATTTGTGGTTATTTGAACGTTCTCT 4904  
 193 CTGCTGCTTTTGTGTAAGAGGAGAGACAGGTATTAAGCTTCATTAAGCTTTGT 252  
 4905 ATGGGTTAAACAGTCGGGTTGGAATGCTATCACTATCAAAAGAGATTTGGACACA 4964  
 253 ATGGGCTGAAGACGGCCCTAGAGCTGTGCAACAAATTAATATCACATTTCTTAATC 312  
 4965 TTGGTTTACTCAAGTTTACCAATGATGGTTTATTTACATGGAATATGAAGGGAT 5024  
 313 AAGGATTAAGAGAGTAAATGATGCACTTTTATATGAAGAGCTTTGTGATGTG 372  
 5025 CAGTAATATATTAGTTTATATGTGATGATTTCTTATGTTGAAGTTCACAAAAG 5084  
 373 GTTCCTTATATGCTCTTTGATGTGATGACCTCTGATGAACAAATATATCAACAG 432  
 5085 TTATTGATATTTTGTGATCAATTTGAGAGTCAATTTTGAAGTTAAAGTGTGTGTA 5144  
 433 AAGTACATCAACTTATGAGAGATGAAAACCAATTTGAGATGCTAGCTTAGGGGAAA 492  
 5145 TATCAAAATATCTTGTATGATTTGTAACACCAATCTGATATATTTATCTCAAG 5204  
 493 TGAACATATTTCTTGCTTGAAGATGATCAATCTGAGAGTGAATTTTTTGAATCAAG 552  
 5205 AAAAATTTCTCAGAAATTTACTTAAGATTTCAAACTAGATGATCTCATATGGGAAAACA 5264

Db 553 AGAAGATGCTCATGAAGTTTGAAGAAGTTAAATGGAAGCTCAATCTGCTCCA 612  
 QY 5265 TACCTGGATTCCGAATGACAAATATGAAGAAG 5297  
 Db 613 CTCCTTTGGTGGAAATTTGAACCTCTCAAG 645

RESULT 10  
 CG973377 867 bp DNA 1linear GSS 15-DEC-2003  
 LOCUS MBEV307FC mch2 Medicago truncatula genomic clone 35E12, genomic  
 DEFINITION survey sequence.  
 ACCESSION CG973377  
 VERSION CG973377.1 GI:39899156  
 KEYWORDS GSS.  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 867)  
 Town, C.D., Shetty, J., Koo, H. and Feldblum, T.F.  
 TITLE Sequencing of BAC ends from Medicago truncatula  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: MBEV307RC  
 CONTACT: Chris Town  
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel.: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 Seq primer: CAGGAACAGCTATGACC  
 Class: BAC ends.  
 Location/Qualifiers

## FEATURES

1..867  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone\_1fb="35E12"  
 /clone\_1fb="mch2"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Cook, D.R. and Kim, D.J. unpublished"

## ORIGIN

Query Match 1.4%; Score 91.4; DB 29; Length 867;  
 Best Local Similarity 47.5%; Pred. No. 3.4e-06;  
 Matches 272; Conservative 0; Mismatches 301; Indels 0; Gaps 0;  
 QY 4725 TTGCACTATAGATTAATGACAAATAATAGTTGTGATAGGAATGACAAATTCACATT 4784  
 Db 73 TGGATACATTCGAATCTTGATGCGTTGGCAGTCAATTAAGTGAATTTGGCATT 132  
 QY 4785 TAGACGTGAGTCGGCGATCTAATGCTCTATTAATCAATTCACATTTATGTCT 4844  
 Db 133 TAGATTTCAATTCAGCAATCTTAATGGAATCTTGATGAAGAAATTTATGTTGCTCAAC 192  
 QY 4845 TTCTCTCTAATCTGATCTTTGAAGAAAACATTTGTTGTTATTTGAACGTTCTGCT 4904  
 Db 193 CTGCTGGTTTTTGGTAAAGGAGGAGGAGCAAGGTGTAAAGCTTCATTAAGCTTTGT 252  
 QY 4905 ATGGGTTAAACAGTCGGGTTTGGATGCTATCACTATCAAAAGATTTGAAGAGA 4964  
 Db 253 ATGGGCTGAAGACAGCCCTAGAGCTTGTAACAATAATTAATATCACTTTCTTAATC 312  
 QY 4965 TTGCTTTTCAAGTTTACACATATGTTTATTTTCATTTGAATATGAAGAGGAT 5024  
 Db 313 AAGGATTCMAAGAGGAATATGATCAACTCTTATGTAAAGAGCTTTGGATGGTG 372  
 QY 5025 CAGTAATATATTGTTTATATGTTATGATGATATCTTATGTTGGAAGTTCAAAAAG 5084

Db 373 GTTCTTAATGTCTCTTTGTATGTGATGACCTGCTAGTAAACAGCAATATCAACAG 432  
 QY 5085 TTATGATTAATTTTGTGATCAATTTGAGATCATTTTGAAGTTAAGTTGGTGA 5144  
 Db 433 AAGTACATCACTTATGAGAGATGAAAACAGTTTGAATGTCTAGCTTAGGGGAA 492  
 QY 5145 TATCAATTAATCTTTGATTTGAATTTGTTAAACCGAATCTGTTATATTTTATCTAAG 5204  
 Db 493 TGAATATTTCTTTGGGCTTGGAAGTGCATCATCTGAGGTGAATTTTTTGAATCAAG 552  
 QY 5205 AAAATTTCTCAAGAAATTAATTAAGATTTCAACTAGATGATCATATGGAAGAAACA 5264  
 Db 553 AGAAGATGCTCATGAAGTTTGAAGAAGTTTAAATGAAGAGTGAATCTGCTCCA 612  
 QY 5265 TACCTGGATTCCGAATGACAAATATGAAGAAG 5297  
 Db 613 CTCCTTTGGTGGAAATTTGAACCTCTCAAG 645

RESULT 11  
 CG961676 869 bp DNA 1linear GSS 15-DEC-2003  
 LOCUS MBECA35TF mch2 Medicago truncatula genomic clone 24E21, genomic  
 DEFINITION survey sequence.  
 ACCESSION CG961676  
 VERSION CG961676.1 GI:39893571  
 KEYWORDS GSS.  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.

REFERENCE 1 (bases 1 to 869)  
 Town, C.D., Shetty, J., Koo, H. and Feldblum, T.F.  
 TITLE Sequencing of BAC ends from Medicago truncatula  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: MBECA35TR  
 CONTACT: Chris Town  
 TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel.: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 Seq primer: TGTAAACAGACGGCCAGT  
 Class: BAC ends.  
 Location/Qualifiers

## FEATURES

1..869  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone\_1fb="24E21"  
 /note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
 HindIII; Cook, D.R. and Kim, D.J. unpublished"

## ORIGIN

Query Match 1.4%; Score 91.4; DB 29; Length 869;  
 Best Local Similarity 47.5%; Pred. No. 3.4e-06;  
 Matches 272; Conservative 0; Mismatches 301; Indels 0; Gaps 0;  
 QY 4725 TTGCACTATAGATTAATGACAAATAATAGTTGTGATAGGAATGACAAATTCACATT 4784  
 Db 73 TGGATACATTCGAATCTTGATGCGTTGGCAGTCAATTAAGTGAATTTGGCATT 132  
 QY 4785 TAGACGTGAGTCGGCGATCTAATGCTCTATTAATCAATTCACATTTATGTCT 4844  
 Db 133 TAGATTTCAATTCAGCAATCTTAATGGAATCTTGATGAAGAAATTTATGTTGCTCAAC 192  
 QY 4845 TTCTCTCTAATCTGATCTTTGAAGAAAACATTTGTTGTTATTTGAACGTTCTGCT 4904  
 Db 193 CTGCTGGTTTTTGGTAAAGGAGGAGGAGCAAGGTGTAAAGCTTCATTAAGCTTTGT 252

QY 4905 ATGGGTTAAAAACAGCTGGGTTTGGATGTATCACACTATCAAAAGATTTGGAAACA 4964  
 DB 253 ATGGGCTGAACAAGCCCTTACAGCTTGATCAACAAATTAATATCCACTTTCTTAATC 312  
 QY 4965 TTGGTTTACCAAGTTTACACAAATGATGTTTATTTACATTTGAATGAAAGAGGAT 5024  
 DB 313 AAGGATTCAAAAGAGGAGAAATGATCAACTCTTTATGTAAGAGCTTTGGATGGTG 372  
 QY 5025 CAGTAATATATTTAGTTTATATGTATGATGATTTATTTATGTTGAAGTTCACAAAAG 5084  
 DB 373 GTTCTTATATTTCTCTTTTGTATGTATGATGACCTGCTAGTAACAGCAATATCAACAAG 432  
 QY 5085 TTATTTGTAATTTTGTGATCAATTTGAGATCAATTTTGAAGTTAAAGTGTGGTGA 5144  
 DB 433 AAGTACATCAACTATTTGAGAGAGATGAAAAACAGTTTGAAGATGTAGCTTGAAGGAAA 492  
 QY 5145 TATCAAAATTTCTGTTATGTAATTTGTAATTTGTAATTTGTAATTTGTAATTTGTAAT 5204  
 DB 493 TGAACATTTTCTTGGCTTGGAGATGATCAATTTGAGATTTTGTGAATCAAG 552  
 QY 5205 AAAAATTTCTAGAAAATTTACTTAAGGATTTCAAACTAGATGATCAATTTGAGAAAACA 5264  
 DB 553 AGAAGATGCTCATGAAAGTTTGAAGAAGTTAAATGGAAGCTGCAAAATGCTCCAA 612  
 QY 5265 TACCTGTGATTCGAAATGACAAATATGAAAAG 5297  
 DB 613 CTCCTTGTGTGGAATTTGAACTCTCAAGG 645

RESULT 12  
 CG928381 940 bp DNA linear GSS 12-DEC-2003  
 LOCUS MBE1G43TRF meth2 Medicago truncatula genomic clone 61H13, genomic  
 DEFINITION survey sequence.  
 ACCESSION CG928381  
 VERSION CG928381.1 GI:39790015  
 KEYWORDS GSS.  
 SOURCE Medicago truncatula (barrel medic)  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.

REFERENCE 1 (bases 1 to 940)  
 AUTHORS Town, C.D., Shetty, J., Koo, H. and Feldblum, T.F.  
 TITLE Sequencing of BAC ends from Medicago truncatula  
 JOURNAL Unpublished (2003)  
 COMMENT Other GSSs: MBE1G43TRB  
 CONTACT: Chris Town  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdtown@tigr.org  
 Seq primer: TGTAAACGACGCCAGT  
 Class: BAC ends.

FEATURES  
 SOURCE  
 1..940  
 Location/Qualifiers  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="61H13"  
 /clone\_1b="meth2"  
 /note="Vector: pBelobAC11; Site 1: HindIII; Site 2:  
 HindIII; Cook, D.R. and Kim, D.J., unpublished"

## ORIGIN

Query Match 1.4%; Score 89.6; DB 29; Length 940;  
 Best Local Similarity 47.5%; Pred. No. 7e-06;  
 Matches 365; Conservative 0; Mismatches 394; Indels 9; Gaps 3;

QY 4483 TAATGATCCGGATGATTAAGTCAATGATGCTGAAGTAAGAAATTTAGATCAAAAAGA 4542  
 DB 118 TAAACATGATGTTGTGAAGCATGCTATGCAAGTTGATCAAGCTCTTGAAGAAAATCG 177  
 QY 4543 TGTTTACGAAGAAGTTCATTTCCACCGGTTGTAGGCTTATTTATGTTGGTGTACA 4602  
 DB 178 TACTGAAAACCTTGATTTTACACCAAAATGTTAAGCTTATGATGTAGATGATTTA 237  
 QY 4603 TACTGAAAATTTGATTTCTTCMAAGGTGTGTTGCGAAATCACCTGTGTTCATGG 4662  
 DB 238 CAAAGTCAAAATTTATGCT---GATGGACCTATTAGAGATATTAAGCAAGTTGTTGC 294  
 QY 4663 CAACAGCAAAAAGAAAATTTGATTTATGACCTTTTGTGTAGTTCACTGTTATGA 4722  
 DB 295 CAAAGTTTACATCAAAATTTGAGGCTTGTATTTATATCATATTTCTCAAGTTGCTA 354  
 QY 4723 TCTTGACTATATATATTGATGATGATTAATGATTTAGGATGAGATGACATTCACA 4782  
 DB 355 ACTCACACAGTCAACACTTTGATTTGCTTTATCTTCATATTAATTTGACATTTACAT 414  
 QY 4783 TTTAGACGTGAGTGGCGATCTTAATGCTCTATTTACTATTCATTAATTCATTTAT 4842  
 DB 415 ACTTGATGCAATTAATGCTTTCTTCATG---GTCACTTGAAGAGATGTATCATGCT 471  
 QY 4843 CTTCCTCTTAATATAGTACCTTTGAGAAAACCATTTGTTGTTATTTGAACGTTCTGT 4902  
 DB 472 GATTCCTCTGGCATCAACCATCAACAAATCAATGTTGCAAGCTTCAAAAGTCTCT 531  
 QY 4903 CTATGGGTTAAAACAGTGGGTTTGAATGATGATCACTATCAAAAGATTTGGAAGA 4962  
 DB 532 CTATGACTTAATATGCTAGTAGAAGAGTATGAAATGATGATCTTACTCTTTTGGC 591  
 QY 4963 CATTCGTTTACTCAAGTTTTCACATGATGATTTATTTACATTTGAATGAAAGG 5022  
 DB 592 TCAAGCATTTTACAGAGCTGATGATCAATTCCTTGTGTTGCAAGAAATGTAA--TC 648  
 QY 5023 ATCAGTAATATATTTAGGTTTATGTTGATGATGATTTCTTATGTTGGAAGTTCAAAA 5082  
 DB 649 ATCATTTACAGTCTCTTGTTGATGATGATGATGATTTCTGTTGAGATTTCTATAG 708  
 QY 5083 AGTTATGATTAATTTTGTGATCAATTTAGAGATCATTTTGAAGTTAAAGTGTGTGA 5142  
 DB 709 TCAATTTGATCATATCAAGTCAATTTTGTGATTTCTATCAAAACAAAGACTTAGCTA 768  
 QY 5143 AATATCAATTAATCTTGGATTTGATTTGTAATTCGTAACGAACTGGTTATTTATCTCA 5202  
 DB 769 ATTTGAATATCTTTAGGATTTAGAGTTGCTTCATTCAAATTTGGGATTTCTTTGTGCA 828  
 QY 5203 AGAAAATTTCTCAAGAAATTTACTTAAGATTTCAACTAGATGATC 5250  
 DB 829 AAGGAATATTTGTAGACTTCTTGTGATTCAGAAATTTAGAGTTG 876

RESULT 13  
 CG821291/c 803 bp DNA linear GSS 18-NOV-2003  
 LOCUS SOYBA90TV LargeInsertSoybeanGenLib Glycine max genomic clone  
 DEFINITION B47J01.MTP7011, genomic survey sequence.  
 ACCESSION CG821291  
 VERSION CG821291.1 GI:38279854  
 KEYWORDS GSS.  
 SOURCE Glycine max (soybean)  
 ORGANISM Glycine max  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseolaeae;  
 Glycine.  
 REFERENCE 1 (bases 1 to 803)  
 AUTHORS Shultz, J., Meksem, K., Shetty, J., Town, C.D., Koo, H., Potter, J.,  
 Wakefield, K., Zhang, H., Wu, C. and Lightfoot, D.A.  
 TITLE End sequencing of BACs comprising a provisional minimal tiling path  
 from a fingerprint physical map of soybean (Glycine max) cultivar  
 Forrest



JOURNAL  
COMMENT

Unpublished (2003)  
 Other\_GSSs: SOTBA90TH  
 Contact: Chris Town, J. L. Shultz and D. A. Lightfoot  
 The Center of Excellence in Soybean Research, Teaching and  
 Outreach, Southern Illinois University at Carbondale and Plant  
 Genomics, The Institute for Genomic Research  
 Room 176, Ag. Building, Mail Code 4415, Carbondale, IL 62901-4415,  
 USA and 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 618 453 1797  
 Email: g4082@siu.edu, jshultz@siu.edu, cdowen@iagr.org (URL:  
 http://bioinformatics.siu.edu)  
 Clones approximating a minimum tiling path were re-arrayed from the  
 library master plates prior to sequencing.  
 For purposes of clone identification each clone name is a  
 concatenation of the original clone location and its new location  
 in the re-arrayed sequencing plates.  
 Seq primer: GTATAGCACTCACTAAGGCGC  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1..803  
 /organism="Glycine max"  
 /mol\_type="genomic DNA"  
 /cultivar="forrest"  
 /db\_xref="taxon:3847"  
 /clone="B4701:MT7011"  
 /note="LargeinsertSoybeanGenLib"  
 /note="Organ: Leaves; Vector: pCUD04541 (pBELOBAC11 EcoRI  
 clones); Site 1: BstXI; Soybean (Glycine max (L.) Merr.)  
 cv. Forrest seeds were grown in greenhouse for fourteen  
 days. Nuclei were isolated and embedded in agarose,  
 restriction digested with Hind III BamHI or EcoRI, large  
 size DNA fragments were ligated in vector V41 (pCUD04541)  
 and electro transformed in DH10a cells. About 90,000  
 clones from BAC libraries were fingerprinted with HindIII  
 and Hae III. Version 2 (automatic build) Contigs were  
 built from 78,001 fingerprints. Contigs were manually  
 examined to find the best non redundant tiling path through  
 the contigs representing about 13,000 clones. The clones  
 were end sequenced."

## ORIGIN

Query Match 1.4%; Score 89.4; DB 29; Length 803;  
 Best Local Similarity 45.9%; Pred. No. 7.8e-06;  
 Matches 343; Conservative 0; Mismatches 401; Indels 3; Gaps 1;  
 4483 TAATGATCCGGATGATTAAGTCAATGATGCTGAATGAGAGAAATTAGATCAAAAG 4542  
 790 TGATCATCCGGAATGGAGACAGCATGATGATAAATACAGGCTTGGAGCATATAG 731  
 4543 TGTTTAGGAAGAAATTCATTTCCACCGGTGGAAGCCATATCTATGGGTGGGTACA 4602  
 730 TACTTGGAGCTAGTCTCTCTCTCCCGAGGAAAGAACTGTGGTGTGGGTGGTCTA 671  
 4603 TACTGAGAAATTAATGATCTCTCAAGGTGTGTTCGGAATCAAGTGTGTCTCCATG 4662  
 670 TGCATATAAAGTTGGGCGCAATAGACAGATTGATAGCCCAAGGCTGATGTAAAC--- 614  
 4663 CAACAGACAAAAGAAAATTTGATTTGACCCCTTTAGTGTATGTTCACTGTATAGA 4722  
 613 CAAGGTTATACCCAGATTATAGACCTAGACATAGAGATACCTTCTCCCTGTGGCTAA 554  
 4723 TCTTGTGACATTAAGATTATGACATATAGGTGGAATGAGATGAGATGACATTAACA 4782  
 553 AATTACTTTTATTTGAGTCTTTCTTGGCATAGCTGCCATCCGTCATTGGGCACTATCA 494  
 4783 TTTAGACGTGAGTGGCGATCTAAATGCCCTCTATTACTCATTTCAATCCAAATTTATGT 4842  
 493 ATTGGAATTTAAATATCATTTCTACTTGAATAATTTGAAGAGAGATTATATGAGCA 434  
 4843 CTTTCCCTCTAAATCAGTACCTTTGAAGAAAACCATGTTGGTTATTTGAACGTTCTGT 4902  
 433 ACCACCGAATTTGTGCTACAGGGGAGTCTAGCTTGTTCCAAACTTGGAGGTCTCT 374

4903 CTATGGGTTAAACAGTCGGGTTTGAATGTATCAACATATCAAAAAGATATTGAGAGA 4862  
 373 TTATGACTCAAAAGATCCCGACAGAGCTTGTTGAAAATTCAGCGCACTGTTCAGGC 314  
 4963 CATTGGTTTATCTCAAGTTTATACAAATATGTTTATTTTCAATGCAATTAAGAAGG 5022  
 313 TTTTGGATGAAAGAAAGTGAAGACCAATTCAGTTTATTTATGTCATACCTCATTTG 254  
 5023 ATCAGTAAATATTTAGTTTATATGTTATGATGATCTTATGTTGGAAGTTACAAAA 5082  
 253 TAGATGTGTTTACTTGTGTGTCTATGATGATGATGTTTTCACAGAAATGAGCAGA 194  
 5083 AGTATGATTAATTTTGTGATCAATGATGATGATGATGATGATGATGATGATGATG 5142  
 193 AAAATGCTCACTGAAGAGACATTTATTTAGCATTTTCACTAAGCATGTGGGAAA 134  
 5143 AATATCAATATCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATG 5202  
 133 ACTTAATATTTTCTTGAATTAAGTTGCTGATGATGATGATGATGATGATGATGATG 74  
 5203 AGAAAAATTTCTCAAGAAATTTACTTAA 5229  
 73 AAGAAAGTATCGTTAGCATCTTAA 47

RESULT 14  
 CG943796 926 bp DNA linear GSS 15-DEC-2003  
 LOCUS MBEMCS77F mbh2 Medicago truncatula genomic clone 84017, genomic  
 DEFINITION survey sequence.  
 ACCESSION CG943796  
 KEYWORDS CG943796.1 GI:39848409  
 SOURCE GSS.  
 ORGANISM Medicago truncatula (barrel medic)  
 Medicago truncatula  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 1 (bases 1 to 926)  
 TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdowen@iagr.org  
 Seq primer: TGTAAACGACGCCAGT  
 Class: BAC ends.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

TIGR  
 9712 Medical Center Drive, Rockville, MD 20850, USA.  
 Tel: 301-838-3523  
 Fax: 301-838-0208  
 Email: cdowen@iagr.org  
 Seq primer: TGTAAACGACGCCAGT  
 Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
 1..926  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /cultivar="genotype A17"  
 /db\_xref="taxon:3880"  
 /clone="84417"  
 /note="Vector: pbeLOBAC11; Site 1: HindIII; Site 2:  
 HindIII; Cook, D.R. and Kim, D.J. unpublished"

## ORIGIN

Query Match 1.4%; Score 89.2; DB 29; Length 926;  
 Best Local Similarity 47.9%; Pred. No. 8.2e-06;  
 Matches 324; Conservative 0; Mismatches 343; Indels 9; Gaps 2;  
 4625 AAAGGTGTTTTCGGAATCAAGTGTGTGTCATGCGCAACAGCAAAAAGAAAATTG 4684  
 18 AAAGTTGCAAGAAACAAAGCAAGGTAGTTGCTCAAGTTACAAACAAAGAAAGTATT 77  
 4685 GATTATGACCTTTTATGTTAGTTACACCTGTATATAGATCTGTGATTAAGATTATG 4744





This Page Blank (uspto)